Final Technical Report

Award G14AP00131

Application of Calibrated Multiple Event Relocation at the National Earthquake Information Center

Dr. Eric A. Bergman
Global Seismological Services
1900 19th St.
Golden, Colorado 80401
Phone (303) 278-4089
bergman@seismo.com

Term of Award:
2014/08/18 - 2015/08/17
(No-cost extension through 2015/10/31)

Report Date:
2016/02/15

Abstract

Under this project I continued an existing research effort to explore the application of calibrated multiple event relocation to the requirements of the National Earthquake Information Center (NEIC) for earthquake monitoring, with the ultimate goal of implementing such a capability in the real-time analysis system, and of supporting the Earthquake Hazards Program of the US Geological Survey. In this phase I worked closely with NEIC personnel to design and implement software and procedures to facilitate import of calibrated earthquake clusters into the COMCAT catalog server in order to provide open internet access to these data. I also developed a data set of 20 calibrated clusters in the central and eastern U.S. (CEUS) that are now being used as input by NEIC research staff in a project to improve the accuracy of the entire existing NEIC catalog of seismicity for the CEUS region.

Introduction

This project is based on a calibrated multiple event relocation analysis that has several advantages over single event location techniques, including the one in use at the NEIC, both in obtaining more accurate locations and also in being able to more accurately estimate the uncertainties of earthquake locations. Specifically, I use the term “calibrated” for earthquake locations that have been determined in such a way that 1) one or more of the hypocentral parameters (typically the epicenter at minimum) can be considered free of the systematic bias that plagues earthquake location research, and 2) that the uncertainties in hypocentral parameters are realistic in a statistical sense, i.e., based on the actual uncertainties of the data themselves and the incorporating the uncertainties of the estimation process. I have been developing this approach, which is based on the Hypocentroidal Decomposition (HD) algorithm introduced by *Jordan and Sverdrup* (1981), for more than 15 years, mainly in the context of nuclear monitoring research. The program which implements this method is called “*mloc*”. The program has been transferred to several NEIC researchers and they have been instructed in its proper application. Training is on-going.

In recent years NEIC personnel have begun to make location accuracy of earthquakes a higher priority than it has been in the past, and I have worked with them for the past several years to explore whether the methods I have developed are suitable for their needs. Under a previously-funded project (*Multiple Event Relocation at the National Earthquake Information Center*, Grant G11AP20016) I conducted calibrated multiple-event relocation analyses in two main contexts, 1) as a quick, but off-line, procedure in support of real-time operations at NEIC, providing verification of preliminary locations upon request, and 2) with NEIC researchers, conducting a detailed analysis of locations of the mainshock-aftershock sequence of the Feb 27, 2010 Mw 8.8 Maule, Chile earthquake (*Hayes et al*., 2013).

These experiences have demonstrated that the ability (given suitable data sets) of *mloc* to calibrated locations of clustered groups of earthquakes is of great value, in particular because of the growing number of NEIC products and services that depend on accurate initial locations from the real-time system (e.g., PAGER). It also became clear during these efforts that catalogs of calibrated locations would have important uses in earthquake hazard reduction research, which of necessity must correlate earthquake activity with the geographic distributions of geological and other geophysical observations, populations, infrastructure and political and social organization. Therefore the emphasis in this project has shifted toward two main goals, one practical in nature and one of a more speculative nature.

The first, more practical, goal addressed under this project is the development of a file format for *mloc* output that can be easily imported to the NEIC's COMCAT catalog server. I also consulted with the COMCAT development team to ensure that calibrated clusters of earthquakes can be routinely published through COMCAT on the internet without losing their distinctive qualities and value, and be made accessible to researchers worldwide.

The more speculative goal of this project is to explore the possibility of leveraging a set of calibrated clusters in a region of interest to obtain more accurate locations for the entire earthquake catalog in that region. In consultation with NEIC personnel, and in part motivated by the remarkable increase in seismic activity in Oklahoma and surrounding states in recent years (*McNamara et al*., 2014) and our previous work together on the August 23, 2011 Mineral, VA earthquake sequence (*McNamara et al*., 2013), I focused on developing a 'framework' of calibrated earthquake clusters in the Central and Eastern United States (CEUS). This framework of 20 calibrated clusters is now being used as input to a research effort by NEIC personnel to derive a full catalog of the CEUS region that has improved location accuracy and better-characterized uncertainties for the hypocentral parameters.

Importing Calibrated Earthquake Clusters into COMCAT

COMCAT ("COMprehensive CATalog server") is an ambitious project at NEIC to implement a publicly-accessible web server that provides access to all NEIC earthquake catalogs and associated data, such as arrival times and earthquake metadata. The traditional catalogs served though COMCAT include historical catalogs and catalogs produced by various researchers and agencies. Therefore they differ in a wide variety of characteristics that must be accommodated, but all of them treat earthquakes as stand-alone entities, i.e., there is no necessary linkage or connection between one earthquake and another.

In serving clusters of earthquakes that have been analyzed as a group, COMCAT faces issues that it was not originally designed to solve. A unique identifier for each cluster, which is carried with the calibrated hypocenter of each member event, is required at minimum, but there are also issues concerned with documenting the location process that has been used for a particular cluster. Even though the same program (*mloc*) may be used for the calibrated relocation of each cluster, the specific procedures used are in general different for different clusters because of the differences in the data set. This issue has been addressed by supplying a commentary for each cluster. Work continues toward formalizing the content of this commentary.

Another issue that required consultation with the COMCAT developers is the need to carry empirically-derived reading errors for individual readings. The use of empirical reading errors (i.e., uncertainty of arrival time readings of a specific phase at a specific station, based on the spread of the actual residuals) is of critical importance in our analysis. It is essential for obtaining unbiased hypocentral parameters because of the utilization of those reading errors in identifying outlier readings, which would cause bias if not removed from a least-squares analysis. The use of empirically-derived readings errors is also necessary in order to derive realistic estimates of uncertainty in those hypocentral parameters. Appendix A is a document supplied to COMCAT developers to describe empirical reading errors and their use in the HD analysis, and it (or a modified version) will be linked in the on-line documentation for every data product served from COMCAT for the calibrated catalog.

Another new parameter for COMCAT arising from this project is the concept of location accuracy codes (also known as “calibration code”). It is extremely useful to employ a simple code that summarizes the level of epicentral accuracy of a calibrated location, and which also carries information about which hypocentral parameters can be considered calibrated. The well-known “GTX” system (e.g., *Bondar et al*., 2004) is inadequate for this purpose. Each event in a calibrated cluster has a calibration code, and in general they are distinct. The document that was provided to COMCAT developers, that will be linked to COMCAT on-line documentation for calibrated data sets, is attached as Appendix B.

A new output file format for *mloc* has been designed and implemented to facilitate import of calibrated clusters into the COMCAT database. Appendix C describes this format.

Calibrated Clusters for the Central and Eastern U.S.

Nearly all clustered groups of earthquakes can be relocated in a relative sense, using a variety of methods for multiple-event relocation (including *mloc*), but the ability to perform a calibrated relocation with *mloc* depends on certain aspects of the arrival time dataset that are difficult to predict without actually attempting the relocation. Nevertheless there are ways to facilitate the search for suitable clusters.

In searching for such clusters in the central and eastern U.S. (CEUS), I began by making a seismicity plot of the region of interest using the catalog of the International Seismological Centre (Figure 1), which has the most complete standard catalog for older events, and especially noting the locations of earthquakes larger than magnitude 4.6. These larger events will have the most relevance to earthquake hazards research and are also likely to be the best recorded events in the catalog. Also regions where larger events occur are likely to be better instrumented so that there may be recordings at shorter epicentral distances that are critical for the location calibration process.

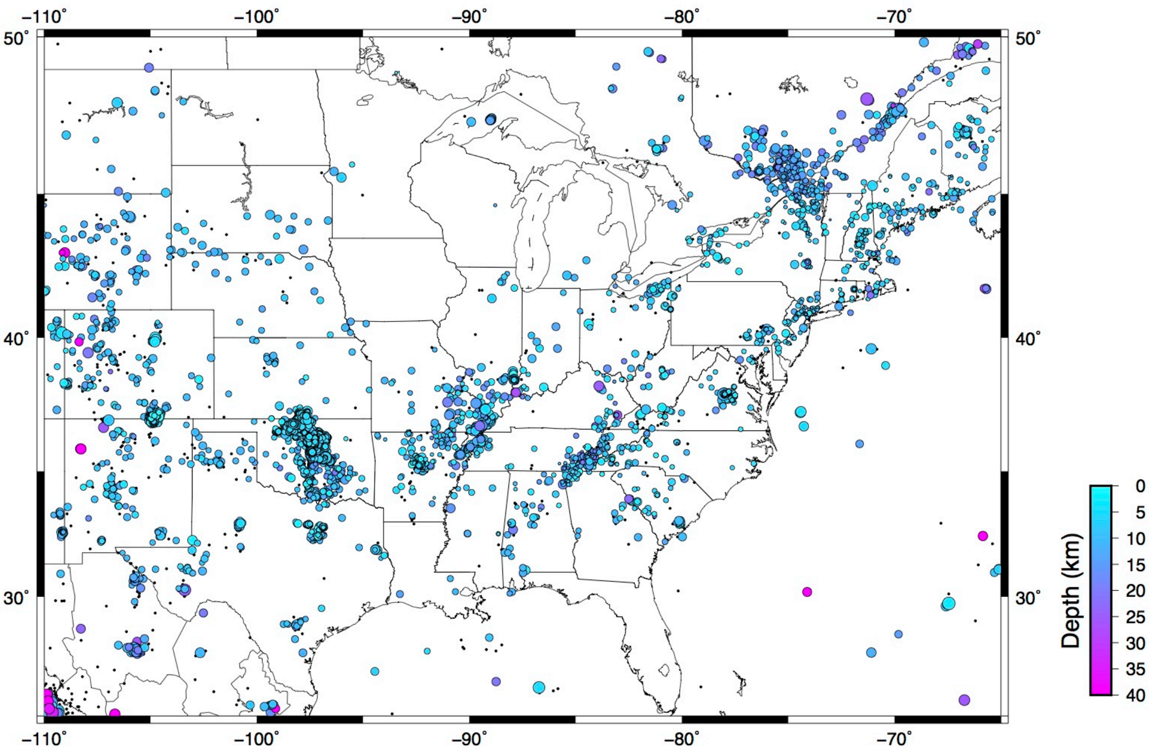

Figure 1. Overview of seismicity of the central and eastern U.S.

For those areas that have been seismically active in the last few decades and contain larger events I formed test clusters by collecting all events within a certain distance (say, 50-75 km) of the largest event. For many of these candidate clusters a simple inspection of the data files was adequate to determine if the data needed for a calibrated relocation was present. A few preliminary relocations were performed if there was still doubt, and these quickly revealed the suitability of a data set for calibrated relocation.

In exploring the very large study area I tried to balance geographic coverage with a reasonable number of clusters. In the long run, of course, it is desirable to carry out calibrated relocations wherever it is feasible, but considering the limited time frame of the current project, it was judged that 20 or so calibrated clusters would be adequate as a trial data set. The locations of these 20 calibrated clusters are shown in Figure 2.

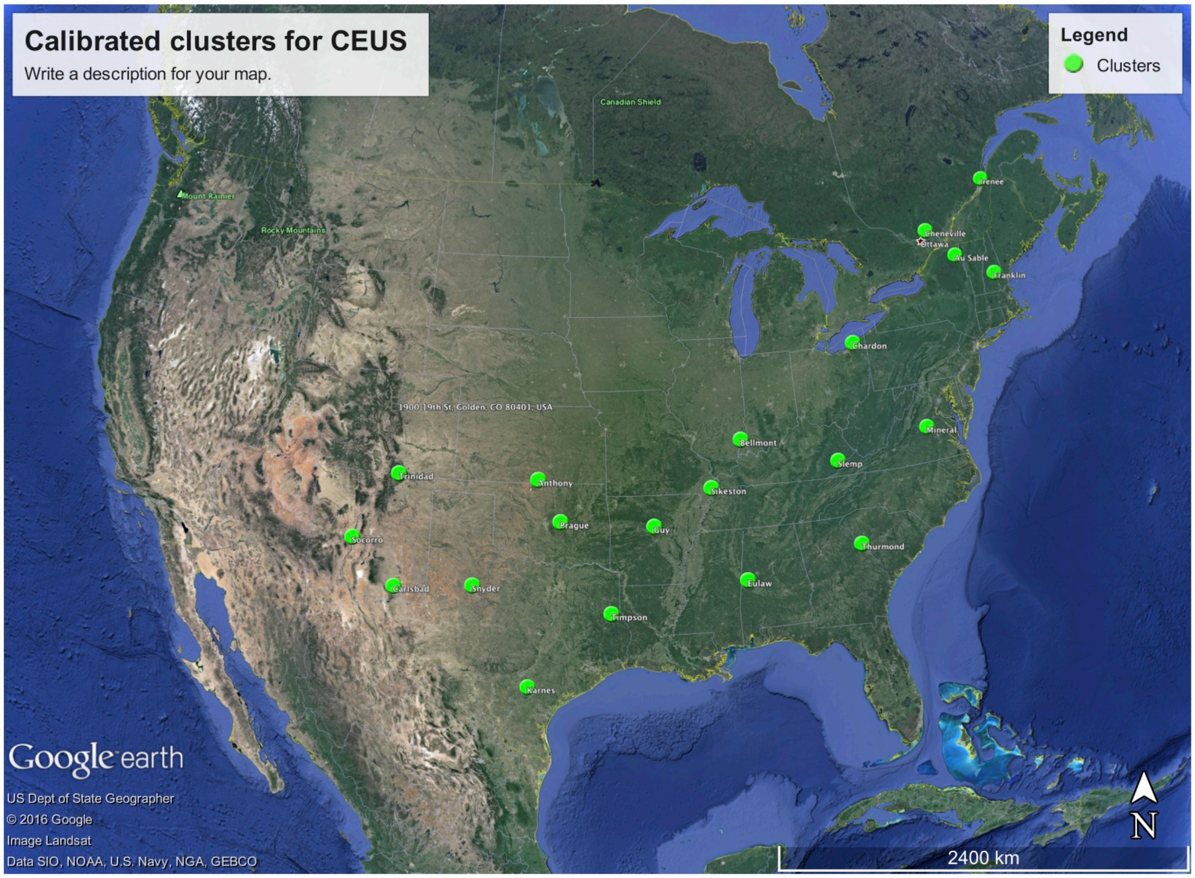

Figure 2. Locations of 20 calibrated earthquake clusters developed in this project to provide a framework for further study.

A short summary of each of the clusters shown in Figure 2 is given in Appendix D, in the form of the 'commentary' for that cluster that was included in the *mloc* output file imported into COMCAT. Each cluster is identified by a name derived from the local geography.

The Next Step: A Fully Calibrated Earthquake Catalog for the Central and Eastern U.S.

The output files formatted for COMCAT from the clusters described above were passed to NEIC personnel at the end of this project, and they are being used as prior constraints in a relocation analysis of the entire catalog of earthquakes in the region, using the program *BayesLoc* (*Myers et al*., 2007). *BayesLoc* has been designed to take advantage of prior knowledge about the locations of some events in a catalog, with the goal of using that knowledge to remove bias in the rest of the catalog. The first results of this attempt have recently been presented in a seminar at the NEIC by one of the researchers (*Yeck*, 2016). The abstract for this presentation is reproduced here:

> "The Central and Eastern United States (CEUS) earthquake catalog reflects decades of variation in seismic networks, location algorithms, and assumed Earth velocity models. These variations manifest themselves as unconstrained biases and uncertainties in earthquake locations. In this study, we standardize the CEUS catalog (~14,000 seismic events) by relocating the catalog in a multiple-event framework. We employ two multiple-event relocation methods. First, where possible, we employ Hypocentroidal Decomposition (HD) to 20 dense, well instrumented, seismicity clusters throughout the CEUS to form a set of calibrated (minimally biased) earthquake clusters. We then leverage these calibrated studies to constrain priors and relocate the full catalog following a Bayesian approach. We employ the bayesloc algorithm which is well suited to this dataset due to its ability to: (1) handle large data sets; (2) constrain parameter error; (3) incorporate a priori hypocenter information; and, (4) model phase assignment errors.
>
> We (1) evaluate our ability to improve our location estimations, (2) evaluate how location quality has varied temporally and spatially, (3) use improved locations and corrections in our travel-time models to evaluate Earth structure, and (4), and examine improved locations of historic large magnitude earthquakes. Improving the accuracy of cataloged earthquake locations parameters (hypocenter locations + origin time) and their associated errors can not only improve our understanding of Earth structure, but also can have a fundamental impact on subsequent seismic hazard analysis and tomographic inversion."

This is a very preliminary result, but is quite encouraging that the general strategy that was followed in this project has potential for systematically and substantially improving the existing earthquake catalog at NEIC that underpins most research on earthquake hazards in the U.S. It has also provided encouragement that the real-time monitoring system at NEIC can be upgraded to provide preliminary locations that are calibrated at a useful level. This will be a subject of future investigations.

Acknowledgements

This project benefitted substantially from collaborations with several persons at NEIC, including Harley Benz, Daniel McNamara, Will Yeck and Michael Hearne.

References

Bondar, I. K., Myers, S. C., Engdahl, E. R., & Bergman, E. A. (2004). Epicentre accuracy based on seismic network criteria. *Geophysical Journal International*, *156*, 483–496. http://doi.org/10.1111/j.1365-246X.2004.02070.x

Hayes, G. P., Bergman, E. A., Johnson, K., Benz, H. M., & Brown, L. (2013). Seismotectonic framework of the 2010 February 27 M_W 8.8 Maule, Chile earthquake sequence. *Geophysical Journal International*. doi:10.1093/gji/ggt238

Jordan, T.H. and K.A. Sverdrup, 1981. Teleseismic location techniques and their application to earthquake clusters in the South-central Pacific, *Bull. Seism. Soc. Am.*, **71**, 1105-1130.

McNamara, D. E., Benz, H. M., Herrmann, R. B., Bergman, E. A., Earle, P. S., Meltzer, A., et al. (2013). The Mw 5.8 Mineral, Virginia, Earthquake of August 2011 and Aftershock Sequence: Constraints on Earthquake Source Parameters and Fault Geometry. *Bulletin of the Seismological Society of America*. doi:10.1785/0120130058

McNamara, D. E., Benz, H. M., Herrmann, R. B., Bergman, E. A., Baldwin, R., & Earle, P. S. (in review 2014). The November 2011 Oklahoma earthquake and aftershock sequence: Constraints on source parameters, fault geometry and aftershock characteristics. *Bulletin of the Seismological Society of America*, 1–35.

Myers, S. C., Johannesson, G., & Hanley, W. (2007). A Bayesian hierarchical method for multiple-event seismic location. *Geophysical Journal International*.

Yeck, W., 2016. Calibrated Multiple-Event Relocations of the Central and Eastern United States, GHSC Seminar presented in Golden, Colorado on February 4, 2016.

Appendix A

Empirical Reading Error

The concept of "reading error" in earthquake location is normally understood as an estimate of the uncertainty of the reading of the arrival time ("pick") of a specific seismic phase on the seismogram of a specific earthquake. Seismic analysts rarely provide their own estimate of that uncertainty beyond a qualitative characterization as "emergent" or "impulsive", and earthquake location codes that employ a quantitative estimate of reading error, e.g., for inverse weighting, normally use an *ad hoc* value based on phase type. It is very important to realize that estimates of hypocenter uncertainty in any earthquake location algorithm depend on the accuracy of the uncertainties assumed for the data, as well as the proper treatment of other sources of uncertainty in the estimation process. This is a significant weakness of most earthquake location analyses.

“Empirical reading error” is a related concept, based on multiple event relocation, i.e., location analysis of a clustered group of earthquakes. The specific implementation discussed here is the one developed for use in the program *mloc* that is based on the Hypocentroidal Decomposition (HD) algorithm of Jordan and Sverdrup (1981). Seismic stations often observe the same seismic phase for multiple events in such a cluster. The resulting multiple observations of the same "station-phase" provide an opportunity to carry out a statistical analysis which leads to an estimate of the uncertainty of those readings that is based on the readings themselves, thus "empirical". It would be more correct to refer to this as an "empirical reading uncertainty" or even “empirical reading consistency”, but we follow the traditional seismological terminology. It is also important to note that this concept of empirical reading error includes contributions to the scatter of readings beyond reading error *per se, i.e.*, i.e., the ability of the analyst to specify the “correct” time of onset time of a seismic phase arrival. For example it also absorbs differences in travel time to a station through a heterogeneous Earth from events that are not exactly co-located, as well as scatter arising from the different philosophies of arrival time picking used by different analysts, differences caused by picking from different channels or instrument responses, changes in station equipment, minor changes in instrument location, irregularities in timing systems, differences in the precision to which picks are reported, etc.

Empirical reading errors are estimated as the spread of the distribution of travel time residuals for a given station-phase (for example, the Pn phase at station TUC) for a specific cluster of earthquakes whose differences in location are typically, but not necessarily, small compared to the separation of the cluster and the station. The number of samples can range from two to several hundred. The analysis is done on the set of residuals obtained by subtracting a theoretical arrival time, based on some travel time model and the current hypocenter of the event, from each arrival time observation. Thus each arrival time reading of a given station-phase is assigned the same empirical reading error. Although this obviously falls short of the ideal of having a reliable estimate of the uncertainty of each reading, it is a significant improvement over the traditional methods for handling uncertainties in arrival time data. Because the arrival time readings are weighted inversely to their empirical reading errors in the location algorithm, the specification of reading errors has a major impact on the estimated hypocenters and their uncertainties.

The estimate of spread of the residuals must be done with a robust estimator, i.e., one that is not sensitive to outliers, which are very common in arrival time data sets. The familiar statistic *standard deviation* is a very non-robust measure of spread. We employ the estimator *Sn* proposed by Croux and Rousseeuw (1992). Note that this has nothing to do with the seismic phase Sn. This measure of scale or spread has three desirable properties, 1) it requires no assumptions about the nature of the underlying distribution, 2) it requires no estimate of the central tendency (e.g., the mean or median) of the distribution, and 3) it reduces to the standard deviation if applied to a Gaussian distribution.

An important aspect of the relocation process consists of multiple cycles in which the current estimates of empirical reading error are used to identify outlier readings, which are then flagged so that they will not be used in subsequent relocations. In the following relocation, the new estimates of empirical reading errors will tend to be smaller because of the filtering of outliers and improvement in the locations of the clustered events. Therefore the process of identifying outliers is iterative and it must be repeated until convergence. In this context, "convergence" means that the distribution of residuals for a given station-phase is consistent with the current estimate of spread. As outlier readings are flagged, the distribution is expected to evolve toward a normal distribution with standard deviation σ equal to the empirical reading error. We generally continue this "cleaning" process until all readings used in the relocation are within 3σ of the mean for that station-phase, where σ is the current estimate of empirical reading error for the relevant station-phase. The procedures used to construct confidence ellipses and other estimates of hypocentral parameter uncertainty in *mloc* (and most other location algorithms) are based on the assumption that the residuals have a normal distribution. In the presence of outlier readings, the resulting uncertainty estimates will be biased.

Because of inverse weighting in the HD algorithm, it is necessary to impose minimum allowed values for empirical reading errors to prevent unrealistically small estimates from arising when the sample size is small and values are very near one another, or identical. We normally use 0.15 s as a minimum which is generally appropriate for the data sets obtained from standard earthquake catalogs, but smaller values can be permitted in special circumstances. For singlet station-phase arrival time data (only one observation) default values that are typical of many earthquake location algorithms (e.g., 0.5 s for teleseismic P) are applied. Singlet readings make no contribution to the estimate of relative locations in the HD algorithm, but they can be used to estimate the hypocentroid, in which case the reasonableness of the default value of empirical reading error must be evaluated for the particular data set in order to have confidence in the derived hypocentral parameter uncertainties.

In summary, the use of empirical reading errors in *mloc* allows us to treat the derived hypocentral uncertainties with considerable confidence. Although any multiple event relocation method could implement a similar analysis, we are not aware of any that do. Single event location methods are inherently handicapped by the lack of any way to investigate data uncertainty in a statistically robust way, although careful attention to the arrival time picking process can partially compensate. Failure to adequately characterize data uncertainties in the hypocenter estimation process leads to bias in the derived parameters and their uncertainties.

References

Croux, C., & Rousseeuw, P. J. (1992). Time-efficient algorithms for two highly robust estimators of scale. *Computational Statistics*, *1*, 411–428.

Jordan, T. H., & Sverdrup, K. A. (1981). Teleseismic location techniques and their application to earthquake clusters in the South-Central Pacific. *Bulletin of the Seismological Society of America*, *71*(4), 1105–1130.

Appendix B

Earthquake Location Accuracy Codes

This document provides a description of a system of codes used to characterize the accuracy of earthquake locations determined with *mloc*, a multiple event relocation program based on the Hypocentroidal Decomposition algorithm of Jordan and Sverdrup (1981), but extensively developed for application in calibrated relocation studies, i.e., relocation studies that are specialized to provide minimally biased estimates of hypocentral parameters and realistic estimates of their uncertainties.

The classification system described here is an extension of the well-known “GTX” system (e.g., Bondar et al., 2004). The primary extension is to generalize the single “class” of the GTX system to four classes that allow an accuracy code to be assigned to any hypocenter. A great advantage of this extra complexity is the ability to distinguish between the different ways in which constraints on location accuracy may have been derived. Moreover we extend the GTX system to carry information about the accuracy of the hypocentral parameters focal depth and origin time, rather than the epicenter alone. This new classification system takes its name “GCNU” from the first letters of the names of the four classes:

- G: ground truth
- C: calibrated
- N: network geometry criteria
- U: uncalibrated

The general form of a location accuracy code in the GCNU system is four characters, of which the first is one of the letters indicating accuracy class, as listed above. The second character carries information on which hypocentral parameters can be considered calibrated. The third and fourth characters are numeric and together provide a length scale in km for the accuracy of the epicenter (equivalent to the “X” term in the GTX system. There are several exceptions to these general rules, as noted below.

Ground Truth: the G Class

This class has only two instances, both of which have only three characters. The GT0 nomenclature is reserved for traditional (or literal) ground truth, events for which all four hypocenter coordinates are known *a priori* at levels of accuracy which are negligible for the purpose at hand. For epicenter and focal depth these uncertainties are typically less than about a hundred meters. At a typical crustal P velocity of 6 km/s 100 meters represents 0.015 s, so origin time should be known to several hundredths of a second in order to be compatible. These limits may not be suitable for some engineering purposes or specialized source studies. The designation "ground truth" has traditionally been reserved for nuclear tests and carefully engineered chemical explosions. It is possible to obtain this level of accuracy with natural seismic sources that are especially heavily instrumented at close range but it is still preferable to use the C class in such cases.

There is a need for a somewhat relaxed ground truth category, because even though the hypocentral parameters of a man-made explosion may be given *a priori*, the uncertainties may not meet the stricter requirements given above. This may be the case because of inadequate record keeping or the difficulty in carrying out suitably accurate surveying or timing prior to the availability of GPS technology. The GT1 category is meant for such cases. This still implies near-certain knowledge of location within a kilometer or so, with comparable uncertainty in origin time (several tenths of a second). Industrial explosions and even some nuclear tests may not meet this standard. Such events ought to be treated in the calibrated (C) class of events, as discussed below, rather than being assigned ground truth status with inflated scale lengths.

No length scale greater than 1 should be used in this class. If the uncertainty is greater than that it is not ground truth.

Calibrated Events: the C Class

In contract to the ground truth class, where the concern is primarily with the scale of random error in the hypocentral parameters, the class of calibrated events is dominated by concern that the estimation process which has been used to determine hypocentral parameters may have introduced significant bias. Therefore we are very much concerned about minimizing bias and understanding which hypocentral parameters may be treated as effectively bias-free. Obviously we also desire to estimate the hypocentral parameters such that the formal uncertainties (driven by uncertainty in the data), usually expressed in Gaussian terms, are as small as possible; this will be handled similarly to the "X" in the GTX formulation, discussed below in the section "Scale Length".

A very important point about the calibrated class of events, is that it includes only events for which the epicenter (at least) has been determined in such a way as to minimize bias. Although a bit unsatisfying in a logical sense, this policy reflects the reality that the seismological community overwhelmingly thinks of ground truth or GT events (using the popular current nomenclature) as referring only to the epicenter. The other important point to be made is that this class requires an actual location analysis, not just the application of some set of network geometry criteria such as those presented by Bondar et al. (2004). In other words, application of network geometry criteria to estimate location accuracy may be a precursor to calibration analysis, but is not a substitute for it.

Given that we do not know the Earth's velocity structure to sufficient accuracy, the only way to reduce bias for an event that was not engineered is to keep path lengths through the unknown Earth structure as short as possible. In other words only near-source data should be employed for estimating calibrated parameters. "Near source data" is not restricted to seismological stations at short epicentral distance, although that is by far the most common case. Mapped surface faulting, treated with all due geological sensitivity, may serve as near source data for the purpose of constraining an epicenter, as may InSAR or other types of remote sensing analyses, since the ultimate signal (e.g., surface deformation) is not subject to bias from unknown Earth structure. InSAR analysis, through determination of the distribution of rupture on a fault plane, may be used to reduce bias in focal depth. Waveform modeling (even at regional or teleseismic distances) may similarly provide useful near-source constraint on focal depth through analysis of the interference of direct and near-source surface-reflected phases.

Unfortunately, there is no methodology for obtaining usefully-calibrated hypocenters for deep earthquakes because every available data type must propagate through an excessive volume of material with insufficiently well-known velocity. The exact definition of "deep" in this context must be evaluated on a case-by-case basis, but it probably includes any event deeper than about 100 km. If uncertainties in velocity structure (and their effect on raypath geometry) are honestly propagated into the uncertainties of the derived location parameters, then the issue will be resolved by the increasing uncertainty of the location, leaving aside the question of bias.

The nomenclature for the calibrated class is based on the following practical considerations about the calibration of the various hypocentral parameters:

Epicenter

Bias in epicentral coordinates can be minimized by means of seismological analysis (typically a location analysis), as well as by other means, including geological and remote-sensing analyses and *a priori* knowledge of human-engineered sources that may be too weak for ground truth status. It is quite common for the epicenter to be the only hypocentral parameter of an event that can be usefully constrained with minimal bias.

Depth

Focal depth is more difficult to constrain than the epicentral coordinates. In the location analysis, it requires data at epicentral distances comparable to the focal depth itself, a few tens of kilometers for crustal events, a much stricter requirement than for the epicenter, which can be usefully constrained with stations 100 km or so away. This distance requirement can be ignored for waveform modeling, however, as well as for analyses of teleseismic depth phases, most famously emphasized by the EHB algorithm (Engdahl et al., 1998). Therefore the minimization of bias in focal depth can be part of the general location analysis, coupled with the estimate of a minimally-biased epicenter, or it can be constrained independently, even when the epicenter may be uncalibrated.

Origin Time

Calibration of origin time is only fully possible when both the epicenter and focal depth can be calibrated. Unless it has been specified *a priori* for a human-engineered event it must be estimated from seismic arrival time data at the shortest possible epicentral distances, and any bias in the location parameters would propagate into origin time. It is quite common, however, to encounter cases where the epicenter and origin time of an event can be constrained with near-source data (not necessarily for the event in question but through linkage to other events in a multiple event analysis), but the focal depth of the event cannot be usefully constrained, other than as an average depth for a cluster of events, some of which have well-constrained depths, or through regional seismotectonic considerations. In this case the origin time itself cannot be considered to be unbiased, but since it is reliably coupled to the assumed focal depth, the combined hypocentral coordinates can still provide valuable information on empirical travel times from a specific point in the Earth.

Given the above considerations there are three cases that need to be distinguished in the calibrated class of the nomenclature. In the following table, the asterisk indicates parameters that have been calibrated in some manner:

	Epicenter	Focal Depth	Origin Time
CH	*	*	*
CT	*		*
CF	*	*	
CE	*		

CH ("H" refers to hypocenter). All four hypocentral coordinates have either been inferred by means that yield minimally-biased estimates or constrained *a priori* (as in some human-engineered events that don't quite qualify for GT1 status or better).

CT ("T" refers to travel time). Epicenter has been calibrated; depth has been fixed at some assumed value (e.g., the average depth of nearby events with constrained depths); the estimate of origin time is based on local-distance data, but relative to an uncalibrated depth. Neither the focal depth not origin time can be considered calibrated in themselves but the combination can be used to estimate empirical travel times from the specific point in the Earth. Such events are not quite as valuable as CH events but still have considerable value as input to model-building exercises or as validation events.

CF ("F" refers to focal depth). Epicenter and focal depth have been calibrated, but not origin time. An example could be an inSAR location for an event and depth calibrated either by an additional analysis of surface deformation to infer distributed displacement on a fault surface, or through waveform analysis. The estimate of origin time is not based on near-source readings. These events can be used in validation exercises where their epicenters are compared with locations done with ray-traced travel-times through a model.

CE ("E" refers to epicenter). The epicenter is calibrated. As with the CT class, depth has been fixed at some assumed (albeit reasonable) value. If the calibration of the epicenter has not been based on near-source seismic data (e.g., an InSAR location), the estimate of origin time must be based on regional or teleseismic arrivals and therefore cannot be considered calibrated, nor can it be used for estimate of empirical travel times. These events can be used in validation exercises where their epicenters are compared with locations done with ray-traced travel-times through a model.

Network Geometry Criteria: The N Class

Events in the N class are not considered to be calibrated in the sense defined here, but the arrival time data set has been processed with some network criteria (e.g., Bondar et al. (2004), but others are developing similar criteria for different source regions) based on simple metrics such as number of readings and distribution of reporting stations, in order to provide an estimate of epicentral accuracy that is expected to account for systematic location bias. The assumption here is that 1) the data do not permit a calibration analysis because there are insufficient near-source data, or 2) that such an analysis has simply not yet been done (i.e., a bulletin has simply been scanned for candidate calibration events). If a careful relocation analysis has been done to standards that can arguably justify classification as a calibrated event, the C class should be used.

NE ("E" from epicenter). The epicentral accuracy has been estimated with an appropriate network geometry criteria. Focal depth and origin time are uncalibrated. Many so-called "GT Catalogs" are dominated by events in this category. Requires a scale length.

NF ("F" from focal depth). As NE but focal depth is calibrated. Requires a scale length.

Everything Else: The U Class

All seismic events that do not fit into one of the GT, C or N classifications are considered uncalibrated. That does not mean that none of the hypocentral coordinates are calibrated, only that the epicenter is not considered to be calibrated. The following classifications are defined:

UE (“E” from epicenter). No hypocentral parameters are calibrated but there is a credible estimate of epicentral accuracy from a location analysis (confidence ellipse), leaving aside the question of systematic location bias. Requires a scale length.

UF (“F” from focal depth). As UE, but focal depth is calibrated. The subset of events in the EHB catalog (Engdahl et al., 1998) that carries depth estimates based on analysis of teleseismic depth phases would fall into this category, as would any event that has been the subject of a waveform modeling exercise that solves for focal depth. Requires a scale length.

U (uncalibrated). Simply a dot on a map. No credible information is available on location accuracy, epicentral or otherwise. No scale length is used.

Scale Length

With the exception of the “U” category all classifications should carry a scale length, equivalent to the “X” in the GTX formulation. The ground truth (GT) class categories are defined with specific scale lengths, which refer to the uncertainty in both the epicenter and focal depth.

For the Calibrated (C) and Network Geometry Criteria (N) classes the scale length is related to the uncertainty in epicenter only. For the CH class one would have to refer to a more detailed description of the data set to learn anything quantitative about the uncertainty in focal depth. The scale length is an integer, in kilometers, related to the uncertainty of the epicenter. Network geometry criteria always yield a single value for scale length. For the C class, as discussed above, there is no consensus about how the 2-dimensional uncertainty in an epicenter should be reduced to a single number. Three possibilities that seem reasonable when dealing with an ellipse with semi-minor axis **a** and semi-major axis **b** are:

- Nearest integer to the semi-major axis length of the confidence ellipse: nint(**b**).
- Nearest integer to the average of the two semi-axis lengths: nint((**a**+**b**)/2).
- Nearest integer to the radius of the circle with the same area as the ellipse: sqrt(**ab**).

For a circular confidence region all three methods are equal. As the ellipticity of the confidence region increases, there will be substantial differences between the different scale lengths, but the first method will always yield the largest value. For a confidence ellipse with semi-axis lengths 1 and 5 km, for example, the scale length calculated with the three methods would be 5, 3, and 2 km, respectively. In analyses using *mloc* we use the first method, with the largest estimate of uncertainty.

Scale lengths larger than 9 are permitted, but they have rapidly diminishing value in the current research environment. When the scale length of confidence ellipses moves into double digits, one ought to begin to worry about the legitimacy of the assumptions underlying the statistical analysis. Such events may better characterized by one of the uncalibrated categories.

Confidence Levels

As Bondar et al. (2004) pointed out, it is necessary to specify the confidence level that has been used in determining epicentral uncertainties, e.g., as a subscript in the form "$GT_{90}5$" to indicate that the confidence ellipse was calculated for a 90% confidence level. The concept of scale length is meaningless without it. Over more than a decade since the proposal was made, compliance on this point seems to be casual at best. It is admittedly awkward to include the subscript in computer output, and since the nomenclature is primarily intended to be carried in digital files it may be best to leave it out, but with the recommendation to clarify the issue in accompanying documentation. In the case of analyses done with *mloc*, the standard confidence level is 90%.

References

Bondar, I. K., Myers, S. C., Engdahl, E. R., & Bergman, E. A. (2004). Epicentre accuracy based on seismic network criteria. *Geophysical Journal International*, *156*, 483–496. http://doi.org/10.1111/j.1365-246X.2004.02070.x

Engdahl, E. R., van der Hilst, R. D., & Buland, R. P. (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. Bulletin of the Seismological Society of America, 88(3), 722–743.

Jordan, T.H. and K.A. Sverdrup (1981). Teleseismic location techniques and their application to earthquake clusters in the South-central Pacific, Bull. Seism. Soc. Am., 71, 1105-1130.

Appendix C
Format for ".comcat" output files written by *mloc*

Eric Bergman
Global Seismological Services
July 5, 2015
v1.41

This document describes a variant of the the native data format used in the multiple event relocation program *mloc*. Data files with the native format have the filename suffix ".mnf" and the format will be referred to as MNF for "*mloc* native format".

Version 1.4 of MNF is a special-purpose variant of what would be considered the "standard" version, v1.3.1. It was created to address a specific problem, the creation of an output file for import into the Global Calibrated Earthquake Cluster (GCEC) catalog in the USGS COMCAT catalog server. This version of MNF should only be found in ".comcat" output files created by the "ccat" command in *mloc*. Version 1.4 of MNF contains several new record types and the formatting of some record types is different from that used in v1.3.1 of MNF.

Version 1.41 of MNF is a very slight change from the original v1.4 format.

Changes from v1.4

- In **phase records**, the travel time residual is now carried with two decimal places of precision, instead of one.

Changes from v1.3.1

- A **station coordinates record**, identified by "C" in column 1, is now defined.
- A **layer velocity record**, identified by "L" in column 1, is now defined.
- **Depth records** are not supported in v1.4.
- Only one **hypocenter line** (the preferred one) is permitted, and the usage code is omitted.
- Usage codes are stripped out of **magnitude records**.
- A "+" usage code is defined in **phase records** for readings that were used.
- A special event ID is defined in **event records**.
- The format of **phase records** is substantially changed.

Background

See the documentation of MNF v1.3.1 for a full discussion. Version 1.4 of MNF is only supported in versions of *mloc* from 10.0.7, and only as an output format for "comcat" files, which is controlled by the "ccat" command. Version 1.41 is first supported in v10.2.6 of *mloc*, released on July 5, 2015.

Description and Features

MNF is a fixed format based on the common concept of a small set of distinct record types with different formats, identified by a character flag in the first column. Each record is a single line. Each type of information (e.g., event, hypocenter, magnitude, phase arrival) has a specific record type, and there are a few utility record types. The currently-defined record types and their flags are given in the following table:

Flag	Record Type
B	Bulletin
F	Format version
E	Event
H	Hypocenter
M	Magnitude
P	Phase reading
#	Comment
C	Station Coordinates
S	Stop event
L	Layer velocity
EOF	End of file

Unlike all other record types, which are distinguished by the flag in column 1, the **end-of-file record** (“EOF”) uses columns 1-3; it has no other arguments. It should only be found once, at the end of the .comcat file.

An MNF v1.4 file (or “comcat” file) always starts with a **bulletin record** “B”, and it will carry a descriptive comment if it was created by *mloc*. The **bulletin record** is always followed by a **format record** “F”. If there is commentary describing the most important features of the earthquake cluster and its relocation, which is highly recommended, it will follow the **format record** “F” as a series of **comment records** “#”. If a custom crustal velocity model has been used this section will be followed by a series of **layer velocity records** “L”. Otherwise it should be assumed that the ak135 travel-time model was used for all phases. If any station coordinates have been taken from sources other than the NEIC metadata server or the ISC Station Registry, there will next be a series of **station coordinate records** “C”. This constitutes the header block of a comcat file. This is followed by a set of event blocks. The comcat file is terminated by an **end of file record** “EOF”.

Data for a single event is carried in a block of records that must start with an **event record** “E” and end with a **stop event record** “S”. Within the block, data is carried in a combination of **hypocenter** “H”, **magnitude** “M”, and **phase reading** “P” records. Only one **hypocenter**

record "H" is permitted. **Magnitude records** are optional and there is no limit to how many can be supplied.

Magnitudes

Magnitude estimates are carried in a **magnitude record**, one magnitude estimate per record. Multiple magnitude estimates can be carried. Magnitudes can be carried to two decimal places.

Optional Fields

All fields defined below will normally be present in a comcat file written by *mloc*. Under certain circumstances the following record types might be absent:

- Comment records, if no commentary text has been provided.
- Layer velocity records, if ak135 was used for all travel-time calculations.
- Station coordinate records, if all stations can be found in the NEIC metedata server or the ISC Station Registry.
- Magnitude records, if an event has no magnitude estimates.

Defined Record Types

Bulletin Record ("B" in column 1)

Column	Description
1:1	**Line format flag "B"**
5:121	**Bulletin description (a117)**

Format Version Record ("F" in column 1)

Column	Description
1:1	**Line format flag "F"**
5:7	**Format version (f3.1)**

Comment Record ("#" in column 1)

Column	Description
1:1	**Line format flag "#"**
2:121	**Comment, optional (a120)**

Layer Velocity Record ("L" in column 1)

Column	**Description**
1:1	**Line format flag "L"**
8:14	**Depth or layer thickness (f7.3)**
20:24	**Vp (f5.3)**
30:34	**Vs (f5.3)**

Note: *mloc* writes the crustal velocity model with two values for each layer, the depth of the upper and lower interface. Internal interface depths are therefore repeated, giving velocities above and below. This accommodates a model in which a layer can have a linear gradient in velocity. Alternatively, a flat-layered model could be defined by with one line per layer, in which the first parameter is interpreted as layer thickness. Depth/layer thickness is given in km. Velocities are given in km/s.

Station Coordinates Record ("C" in column 1)

Column	Description
1:1	**Line format flag "C"**
3:8	**Station (a6)**
11:18	**Latitude (f8.4)**
20:28	**Longitude (f9.4)**
31:36	**Elevation (i6)**

Note: Station elevation is the elevation of the instrument, given in meters, relative to mean sea level (positive or negative).

Event Record ("E" in column 1)

Column	Description
1	**Line format flag "E"**
5:121	**Event ID (a116)**

Note: Event IDs are built up from the prefix "gcec_" (Global Calibrated Earthquake Clusters), the cluster name, and the event number within the cluster.

Hypocenter Record ("H" in column 1)

Column	Description
1	**Line format flag "H"**
5-8	**Year (i4)**
10:11	**Month (i2)**
13:14	**Day (i2)**
16:17	**Hour (i2)**
19:20	**Minute (i2)**
22:26	**Seconds (f5.2)**
28:32	**OT uncertainty (f5.2)**
35:42	**Latitude (f8.4)**
44:52	**Longitude (f9.4)**
54:56	**S_{min} azimuth (i3)**
58:62	**Error ellipse S_{min}, (f5.2)**
64:68	**Error ellipse S_{maj} (f5.2)**

70:74	**Focal Depth (f5.1)**
76:76	**Depth code (a1)**
78:82	**Plus depth uncertainty (f5.1)**
84:88	**Minus depth uncertainty (f5.1)**
90:93	**GTCU (a4)**
95:102	**Author (a8)**
104:121	**Cluster ID (a18)**

Note: Both depth uncertainties are provided as positive numbers. “Plus” depth uncertainty is on the deeper side; “Minus” uncertainty is shallower, and therefore should not be greater than the focal depth in absolute value.

The “GTCU” field carries a four-character code relating to calibration status (I prefer this term to “ground truth” level). It could be the GTX formulation (e.g., Bondar et al., (2004)), but I have developed the GTCU nomenclature to provide much more detailed information on the subject of what hypocentral parameters are considered to be calibrated (i.e., thought to be bias-free). The GTCU nomenclature is documented fully elsewhere.

The “depth code” in column 76 defines the nature of the depth constraint. Standard values are:

Depth Code	Meaning
c	Cluster default depth
d	Teleseismic depth phases
e	Engineered (man-made explosion)
f	Fault model (InSAR, GPS, etc)
l	Local-distance readings
m	*mloc* solution with free depth
n	Near-source station readings
r	Relocation outside *mloc* with free depth
u	Unknown
w	Waveform analysis

Magnitude Record ("M" in column 1)

Column	Description
1	**Line format flag “M”**
5:8	**Magnitude (f4.2)**
10:14	**Magnitude scale (a5)**
16:110	**Author and comments (a95)**

Phase Reading Record ("P" in column 1)

Column	Description
1	**Line format flag “P”**
3	**Usage flag (a1)**
5:10	**Station code (a6)**
12:17	**Epicentral distance (f6.2)**
19:21	**Azimuth, event to station (i3)**
23:30	**Phase name (a8)**
32:35	**Arrival time year (i4)**
37:38	**Arrival time month (i2)**
40:41	**Arrival time day (i2)**
43:44	**Arrival time hour (i2)**
46:47	**Arrival time minute (i2)**
49:54	**Arrival time seconds (f6.3)**
56:57	**Reading precision (i2)**
59:64	**TT residual (f6.2)**
66:70	**Empirical reading error (f5.2)**

Note: The usage flag will be either “+”, meaning the reading was used for either the hypocentroid or cluster vectors or both, “-”, meaning the reading was not used in the relocation for some reason, or “x” meaning that the reading was flagged as an outlier and not used for relocation. It is important to note that the factors controlling how an individual phase reading was used (or not) in an *mloc* analysis are quite complex and there is no information in the v1.4 format concerning why a reading was not used, except for the special case when it has been identified as an outlier. Reference to other *mloc* input and output files is needed to answer that question.

The arrival time seconds field is written with the precision implied by the “reading precision” field.

Stop Event Record ("S" in column 1)

Column	Description
1:1	**Line format flag “S”**

Note: Only the “S” in column 1 is required, but for better readability it is useful to write “STOP” in columns 1:4.

End of File Record ("EOF" in columns 1-3)

Column	Description
1:3	**Line format flag “EOF”**

Appendix D

Summaries of Calibrated Earthquake Clusters in the Central and Eastern U.S.

Each cluster is described by the ‘commentary’ included with the data file for import into COMCAT, plus a figure showing the locations and uncertainties of the locations. The clusters are:

- Anthony, Kansas
- Au Sable, New York
- Bellmont, Illinois
- Carlsbad, New Mexico
- Chardon, Ohio
- Chenéville, Canada
- Eulaw, Alabama
- Franklin, New Hampshire
- Guy, Arkansas
- Irénée, Canada
- Karnes, Texas
- Mineral, Virginia
- Prague, Oklahoma
- Sikeston, Missouri
- Slemp, Kentucky
- Snyder, Texas
- Soccoro, New Mexico
- Thurmond, South Carolina
- Timpson, Texas
- Trinidad, Colorado

Anthony, Kansas

The Anthony cluster is based on a series of small events that started occurring near Anthony, Kansas in March 2014. The USGS arranged to install three temporary stations, which caught several of the events. Unfortunately the temporary stations were all just off to the NE of the seismicity, but other organizations have installed more stations that now provide excellent coverage.

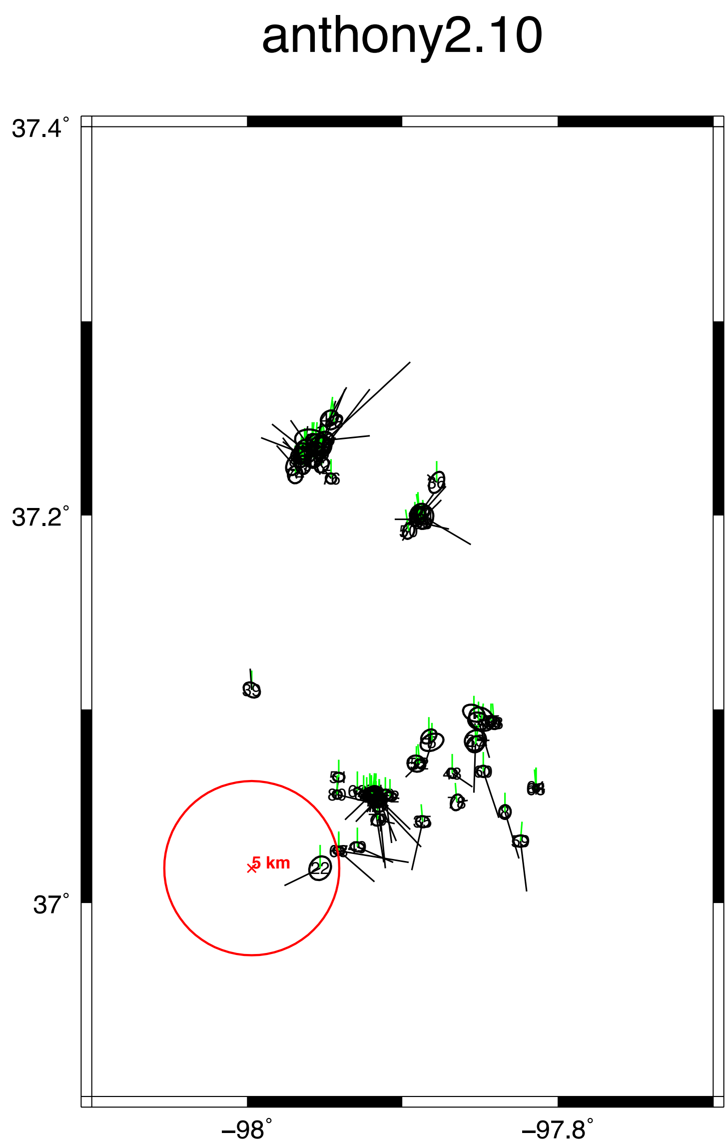

Figure D1. Calibrated locations of the earthquakes in the Anthony, Kansas cluster. Confidence ellipses are for relative location, at the 90% confidence level. The absolute uncertainty of the location of any event also includes a contribution from the uncertainty of the hypocentroid (see text). For scale, a circle of 5 km radius is shown in red. Black vectors indicate the change in location from the standard catalog location. Green vectors indicate the change in location from the starting location for this relocation, which was taken from the previous relocation run.

This version of the Anthony cluster is based on a new search of COMCAT that returned about 1000 events. I selected 88 events that have at least 40 arrival time readings and further trimmed the cluster to 81 events that all have close-in readings for depth control from March 2014 to July 2015. The largest event is magnitude 4.8. This cluster can be relocated with free depth, with most depths concentrated at 6-7 km (4-11 km overall) and uncertainties less than 1 km in depth. Using data out to 1.2° the calibration level is 0.2 km and all events qualify as CE00-01.

Au Sable, New York

The Ausable cluster is named for the Mw 5.3 Au Sable Forks earthquake of April 20, 2002, it's aftershocks and nearby earthquakes. I selected 7 events with close-in readings for depth control and analyzed those first to refine focal depths and velocity structure. I used the crustal model from the cheneville cluster and it worked very well with only very minor adjustments. One event wants to be quite shallow, but the rest are happy in the 8-12 km range. I added other events (the ones with most readings) and got to 20 events in this series. Using data out to 1.2° the azimuthal coverage is excellent and the calibration level is 0.5 km. All events qualify as CE01.

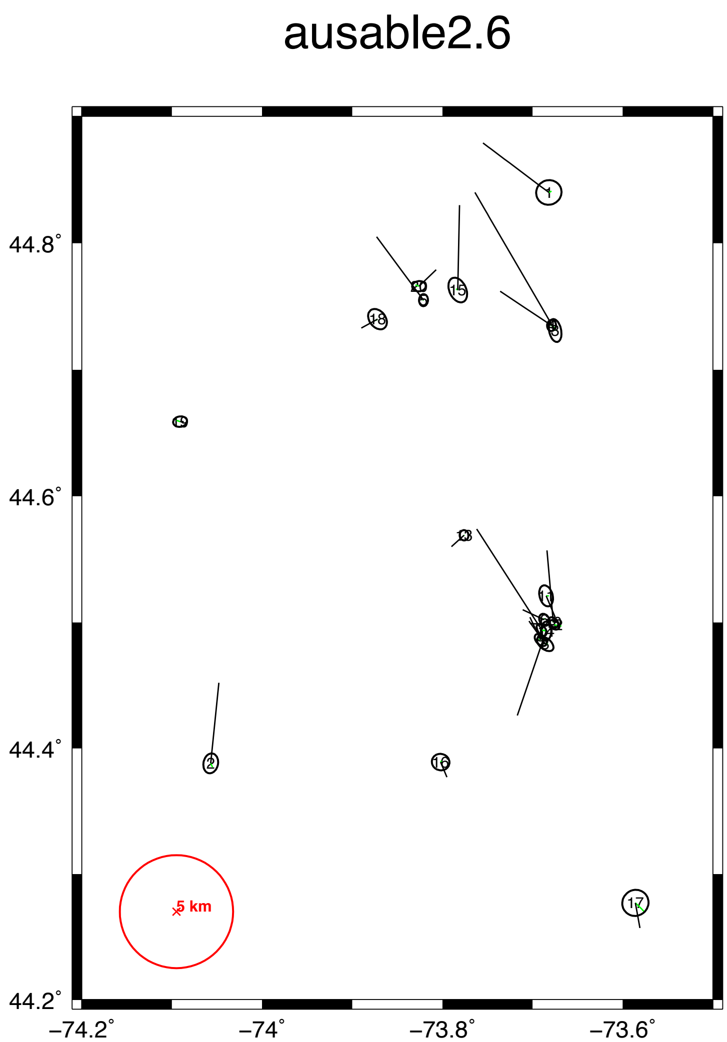

Figure D2. Calibrated locations of the earthquakes in the Au Sable, New York cluster. See Figure D1 for further explanation.

Bellmont, Illinois

This cluster is based on a group of earthquakes just north of the town of Bellmont, Illinois, that includes a 5.0 Mw mainshock on April 18, 2008 and its aftershocks. The cluster contains 25 events, 19 of which are clustered tightly near the town of Bellmont. The other 6 are spread out in different directions, about 50 km from this group. Depths of all events are set manually, most of them from near-source readings, ranging from 17-39 km (most in the range 25-30 km). The 2008 mainshock has a depth of 29 km. Station coverage for direct calibration is very good. I used data out to 1.0° to estimate the hypocentroid and obtained a calibration level of 0.7 km. 24 events qualify as CE01.

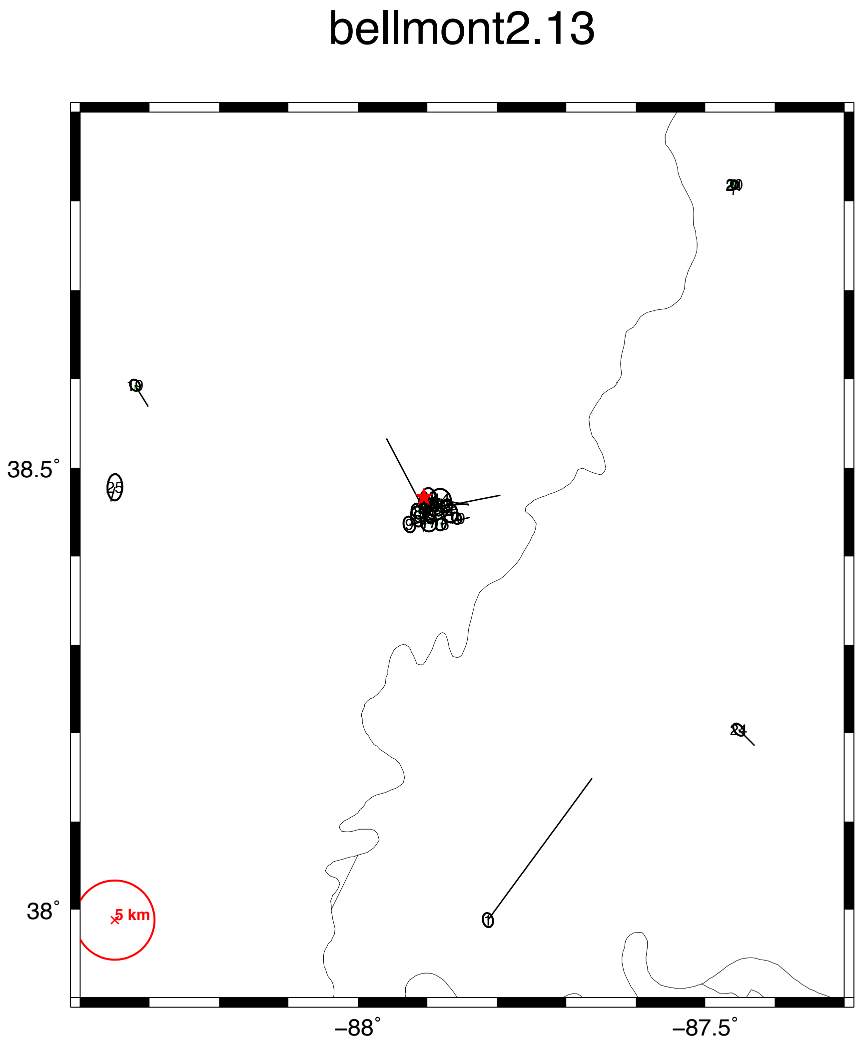

Figure D3. Calibrated locations of the earthquakes in the Bellmont, Illinois cluster. See Figure D1 for further explanation.

Carlsbad, New Mexico

This cluster is named for the nearby city of Carlsbad, New Mexico. It contains 25 events that occurred between 1999-2012, mostly in a limited area with magnitudes up to 4.4. Between local permanent stations and the TA it is well recorded. Direct calibration is done with data out to 1.4°, yielding a calibration level of 0.6°. All but one event qualify as CE01-02. Three events had close-in readings that are used to set focal depths at 7-8 km, and the velocity model was refined with these events. All other events have similar depths, ranging from 5-9 km, when set to be consistent with the ones that have depth constraint.

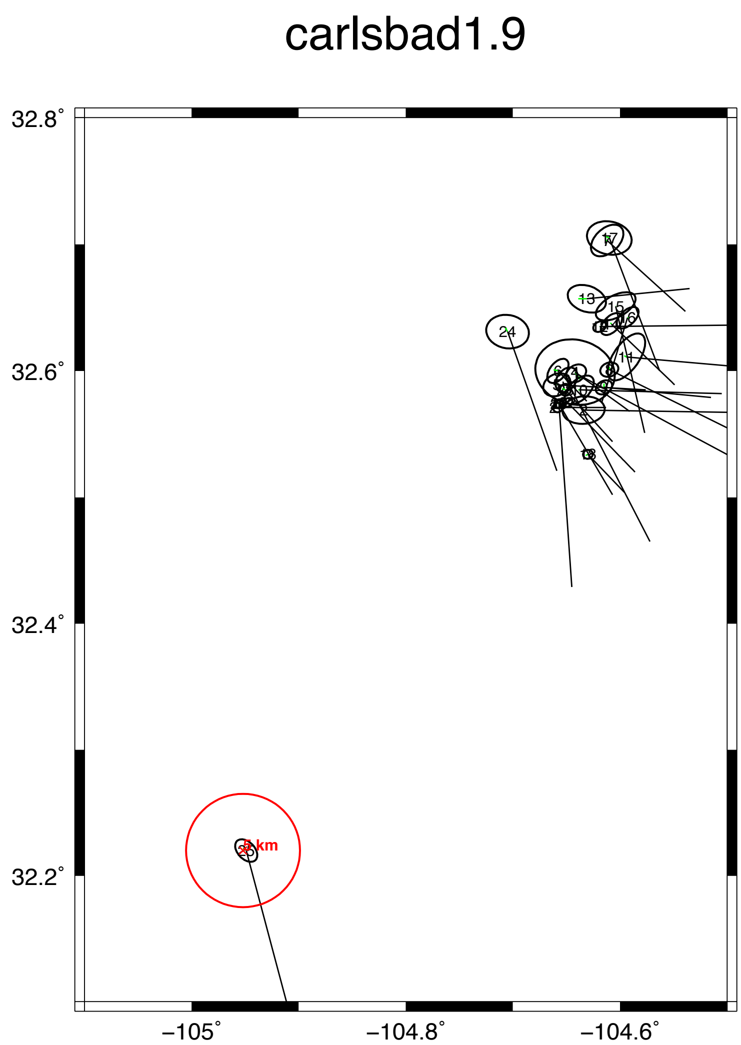

Figure D4. Calibrated locations of the earthquakes in the Carlsberg, New Mexico cluster. See Figure D1 for further explanation.

Chardon, Ohio

The cluster is based on a M4.8 event on January 31, 1986 on the shore of Lake Erie in Ohio. It's named for the nearby city of Chardon, but Mentor is actually closer. The cluster contains 36 events from 1983-2015.

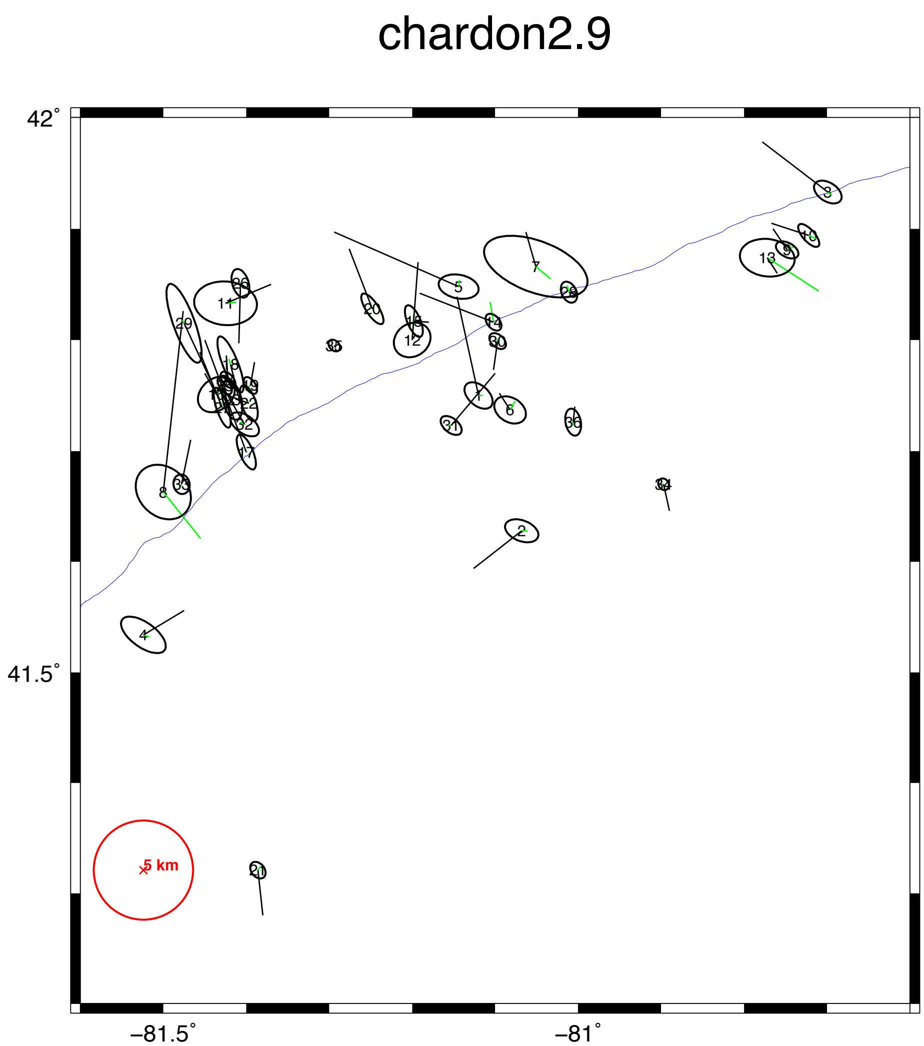

Figure D5. Calibrated locations of the earthquakes in the Chardon, Ohio cluster. See Figure D1 for further explanation.

The cluster is well-calibrated using data out to 1.4°, but there could be problems with bias because the raypaths to the north must cross Lake Erie and all the stations to the north are relatively far away. Focal depth is also problematic for this cluster. Only two events have data at close-enough range to provide direct control on focal depth (11 and 14 km). The velocity model was refined against these two events. About half of the remaining events can be constrained more loosely in depth by considering the fit to the direct and refracted portions of the TT curve, and these have depths from 5-11 km, except for one that seems to be at 22 km. The remaining events were held at 10 km depth for relocation. Calibration level is 0.8 km and 29 events qualify as CE01-02, but there may still be some location bias for the cluster as a whole because of the station geometry issues mentioned above.

Chenéville, Canada

The cheneville cluster is named for the town of Chénéville, northeast of Ottawa, Canada. It consists of 60 events (1987-2014) for which there were readings at stations close enough to constrain focal depth. There are many more events in the area that could be relocated as well, with somewhat poorer constraint on depth. The cluster includes one moderate-sized event (5.4 mb, on June 23, 2010) with an extensive aftershock sequence. Most other events in the cluster range in magnitude from ~2-4.

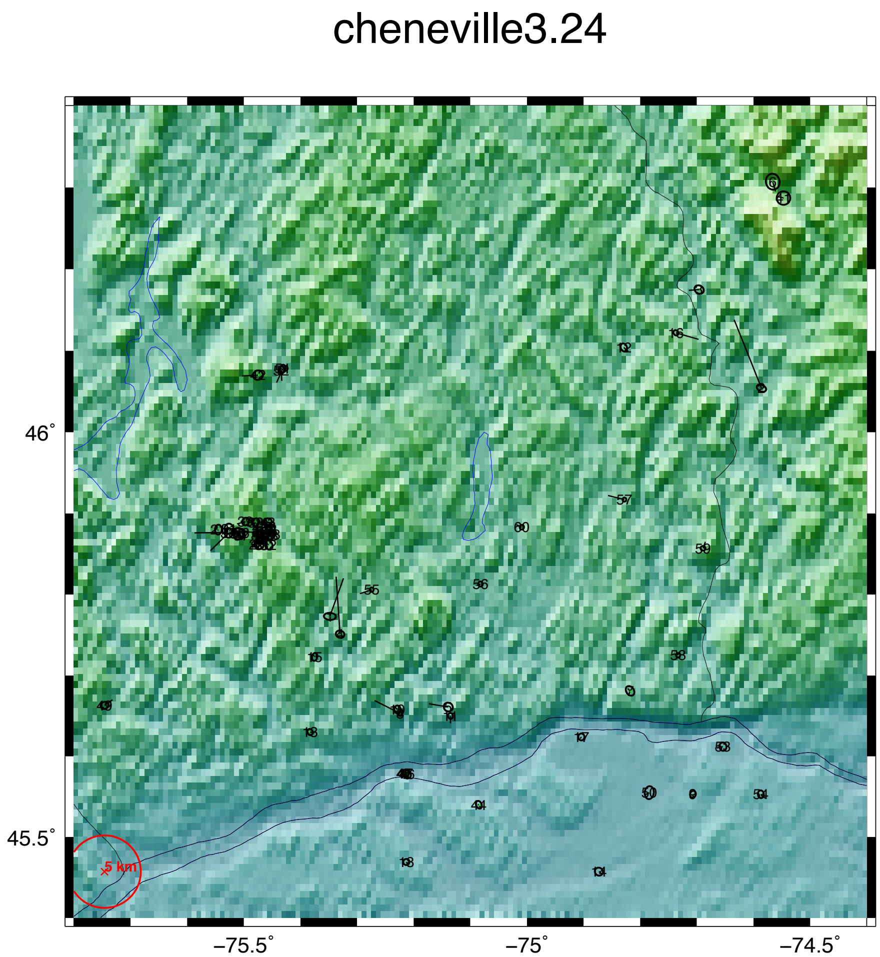

Figure D6. Calibrated locations of the earthquakes in the Cheneville, Canada cluster. See Figure D1 for further explanation.

All arrival time data for this cluster were obtained from the ISC Bulletin. Station coverage is excellent in this area and direct calibration, using over 1000 P and S readings at distances less than 1.2°, provides a calibration level of less than 500 m. The velocity structure was refined from ak135 to fit the observed data and found to be very similar to the model proposed for the

Grenville Province by Viegas et al. (2010) [Regional Wave Propagation in New England and New York. Bulletin of the Seismological Society of America, 100(5A), 2196–2218. http://doi.org/10.1785/0120090223]. Because of the selection of events with near-source readings, it was possible to perform a relocation with focal depth as a free parameter. Focal depths range from 6-21 km with a concentration in the range 12-18 km. Uncertainties in focal depth range from 0.4 to 1.4 km. All events qualify as CE01 for epicentral accuracy.

Eulaw, Alabama

The cluster is named after the nearby town of Eulaw, in Greene Country, Alabama. It is based on a series of small events that started in November 2014 and continues into mid-2015 with magnitudes up to Mw 3.8. The cluster also includes a 4.3 Mw event on November 7, 2004, about 25 km south of the recent seismicity. The cluster contains 18 events. A set of 5 recent events have very good depth control from a temporary seismic station and these were used to refine the velocity structure. The focal depths of remaining events are set to be consistent with these. Depths range from about 5-9 km with uncertainties of several km. Direct calibration, using readings out to 1.0° epicentral distance for the hypocentroid, yields a calibration level of 1.4 km, and 17 out of 18 events qualify for CE02 epicentral accuracy.

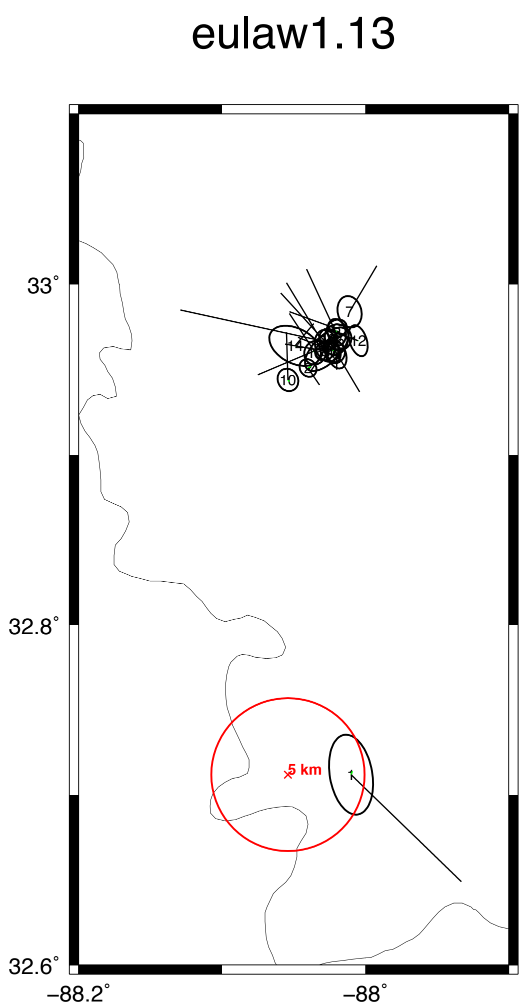

Figure D7. Calibrated locations of the earthquakes in the Eulaw, Alabama cluster. See Figure D1 for further explanation.

Franklin, New Hampshire

This cluster is based on an M4.4 event on January 19, 1982 near Franklin, New Hampshire. The area is pretty active at lower magnitudes and it was not hard to put together a cluster of 32 events (1977-2013) that can be calibrated well. Many of these events are at very shallow depth, 1-4 km and all the rest except one (16 km) are at 8 km or less. I used data out to 0.9° to calibrate this one. I could easily have gone further but the geometry of the network looked a bit unbalanced then. Calibration level is 1.1 km and 28 events qualify as CE01-02.

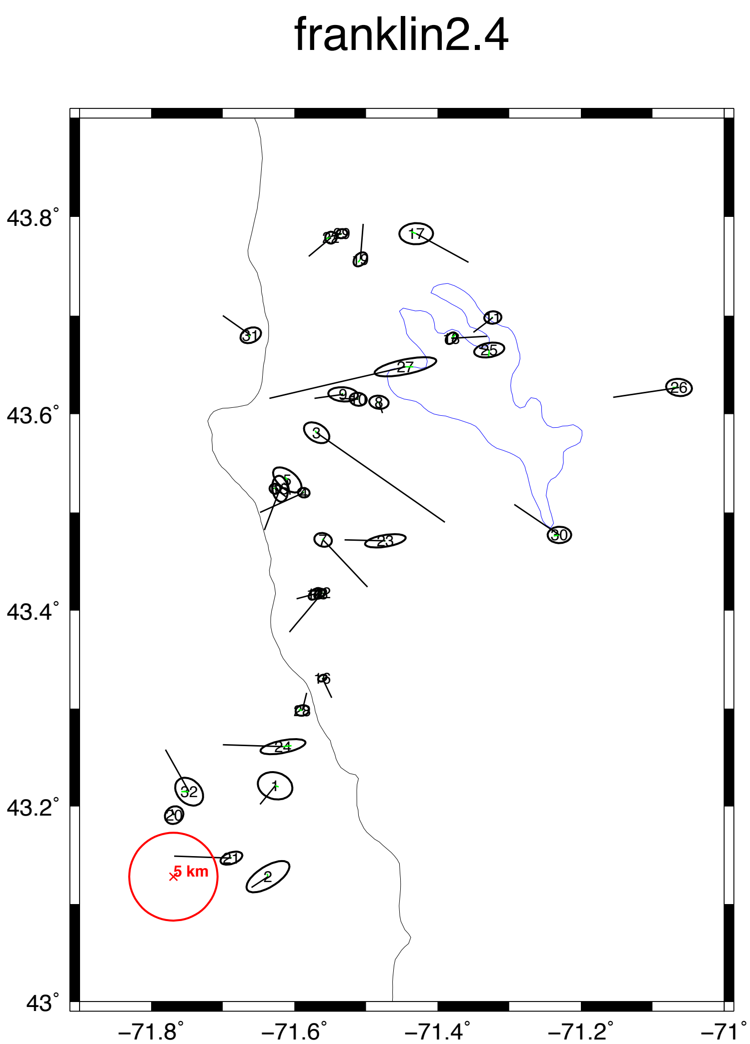

Figure D8. Calibrated locations of the earthquakes in the Franklin, New Hampshire cluster. See Figure D1 for further explanation.

Guy, Arkansas

The Guy cluster is named for the nearby town of Guy, Arkansas. The seismicity appears to be largely related to waste fluid injection, according to Horton, S. (2012) [Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in central Arkansas with potential for damaging earthquake. Seismological Research Letters, 83(2), 250–260. http://doi.org/10.1785/gssrl.83.2.250]. The seismicity is also referred to as the "Conway" earthquakes after a larger town to the south. The data set was pulled from the ISC Bulletin. The cluster is formed of 43 events in the period 2010-2012 that have readings at close range to constrain focal depth but there are several hundred events in the area that could be relocated with good accuracy. The magnitude of these events ranges from 2.4 to 4.5.

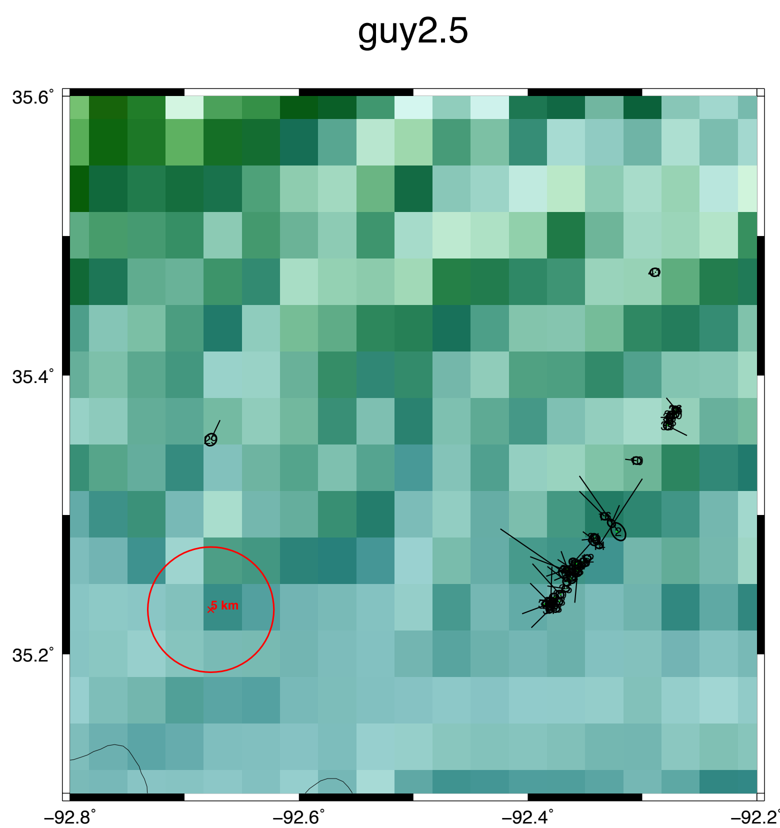

Figure D9. Calibrated locations of the earthquakes in the Guy, Arkansas cluster. See Figure D1 for further explanation.

The crustal velocity model was derived by fitting the observed data, but it is quite close to the model of Chiu et al. (1984) that Horton (2012) used. The relocation could be done with focal depth as a free parameter; focal depths mostly are between 6 and 9 km, with uncertainties of 0.5

to 1.2 km. Using readings out to 0.8° direct calibration yields a calibration level of 0.8 km and all 43 events qualify as CE01 for epicenter accuracy.

Irénée, Canada

The cluster is located in the St. Lawrence Rift System northeast of Quebec, right on the St. Lawrence River. It is named for a small township (Saint-Irénée) in the area. I did a search of the same area in the ISC bulletin (500+ events), first selected 117 events with 40+ readings then trimmed it down to 80 events that had at least two readings with 0.15 degrees for depth control. Largest magnitude is 4.8 on March 6, 2005 and range of dates is 1993-2015. The same crustal model I used for the Cheneville cluster works well here, with just a slight adjustment to Sn velocity and Moho depth. Focal depths cover a rather wide range, from 2-20 km, but the data provide strong constraints on that. I set focal depths by hand; free depth runs were a bit unstable probably because of connectivity problems for the stations controlling depths. Using data out to 1.2° the calibration level is 0.6 km and all but one event qualifies as CE01.

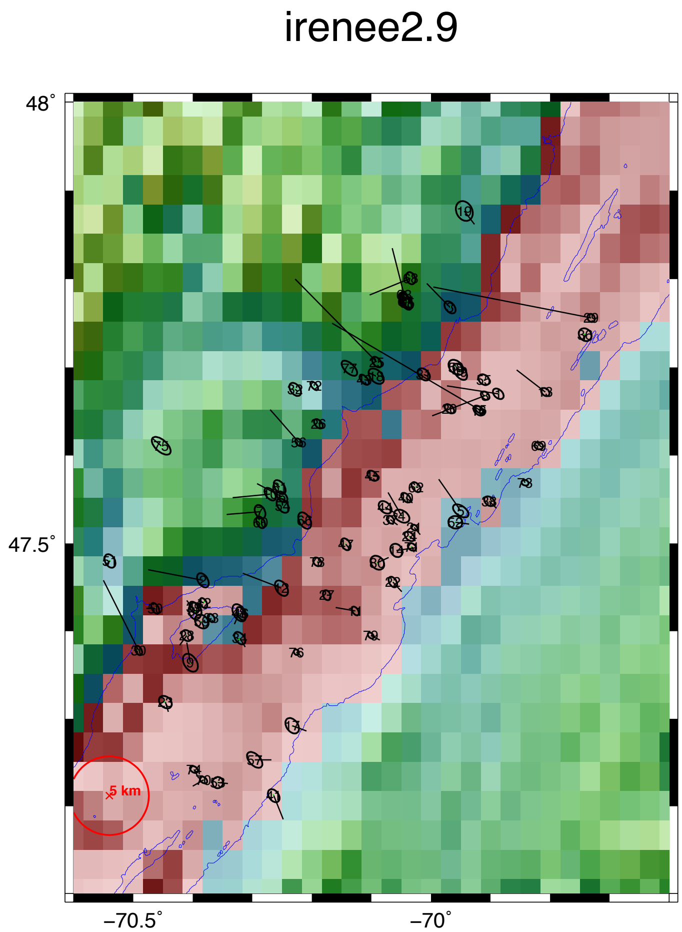

Figure D10. Calibrated locations of the earthquakes in the Irenee, Canada cluster. See Figure D1 for further explanation.

Karnes, Texas

This cluster is named after the nearby town of Karnes in southern Texas. It is based on a M4.8 event on October 20, 2011. There are few other events in the area but I could make a decent cluster of 7 events. Because of the small number of events and the lack of readings at very close distances, there is some tradeoff between focal depths and velocity structure, meaning the focal depths are not well-constrained. In this solution the depths of the four events that have some constraint range from 9-14 km, and the three with little constraint are held at 10 km. Using data out to 1.4° the azimuthal coverage is good; direct calibration gives a calibration level of 1.3 km and all events qualify as CE01-02.

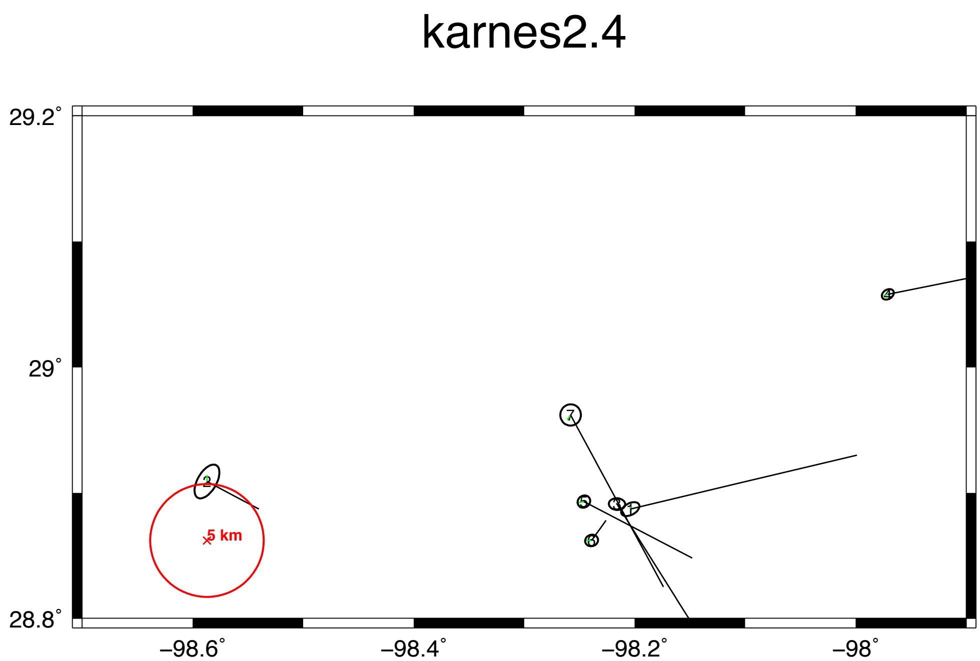

Figure D11. Calibrated locations of the earthquakes in the Karnes, Texas cluster. See Figure D1 for further explanation.

Mineral, Virginia

The mineral cluster is based on the mainshock-aftershock sequence near Mineral, Virginia that began with the 5.8 Mw mainshock on August 23, 2011. The full sequence contains about 450 events but this cluster contains 66 events that have the largest number of arrival time readings for use in location calibration studies. Most of the events have readings from temporary stations that were installed very close to the epicentral area, and they have excellent depth control. Their depths were set in the course of preliminary studies with free-depth relocations that have uncertainties less than 500 m in most cases. These depths range from 2-9 km, concentrated near 6 km. The depths of the mainshock and early aftershocks that do not have close-in readings have been set by matching the pattern of direct and refracted arrivals. Using data out to 0.6° the calibration level is 0.2 km and all events qualify as CE00-01.

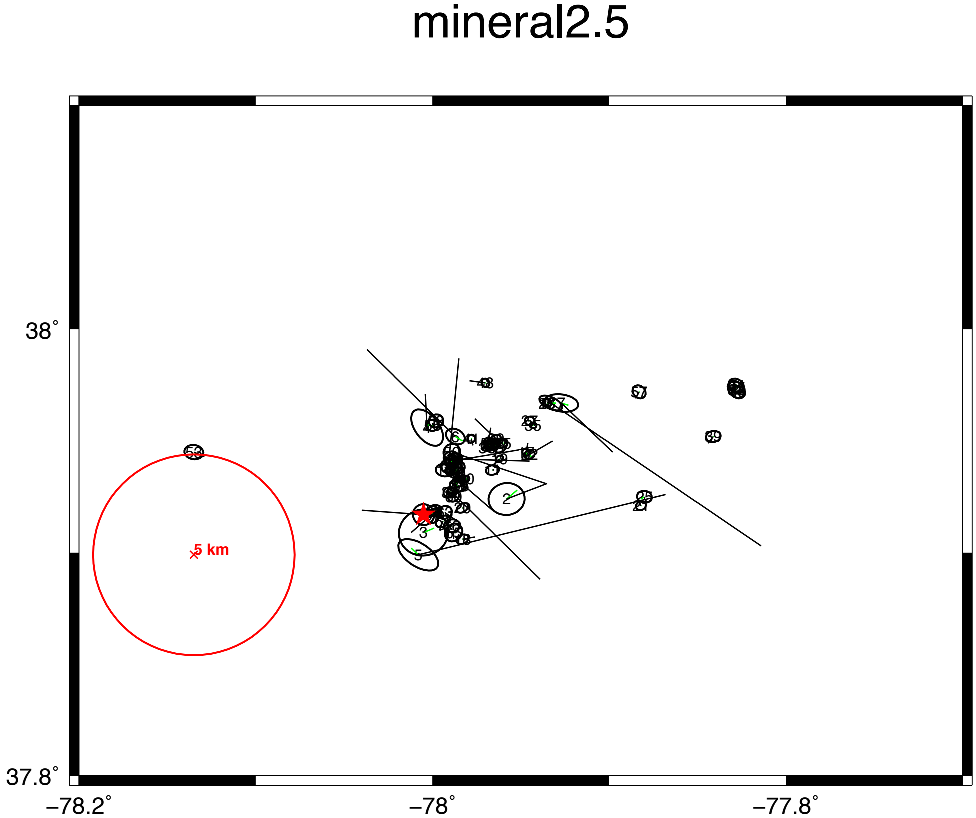

Figure D12. Calibrated locations of the earthquakes in the Mineral, Virginia cluster. See Figure D1 for further explanation.

Prague, Oklahoma

prague16 is a new approach to the Prague, Oklahoma sequence, based on the November 6, 2011 mainshock, 5.6 Mw. I searched COMCAT for events from November 1, 2011 and selected those with 20 or more arrival time readings. Then I added in the readings from our special study of the Prague sequence, many of which had not gotten into COMCAT. The readings from temporary stations provide excellent depth control for aftershocks, beginning about 15 hours after the mainshock. The cluster includes 41 events, beginning with six foreshocks and extending to the end of January 2012, when the temporary stations were removed. About half the events have readings that control focal depth directly, and I started working with these, doing free-depth locations to establish depth and crustal velocity model. Depths range from 5-10 km with uncertainties of 500 m or less. Then the remaining (earlier) events were added to the cluster with depths set manually to match the pattern of direct and refracted arrivals. Using data to 0.8° the calibration level is 0.2 km and all events qualify as CE00-01.

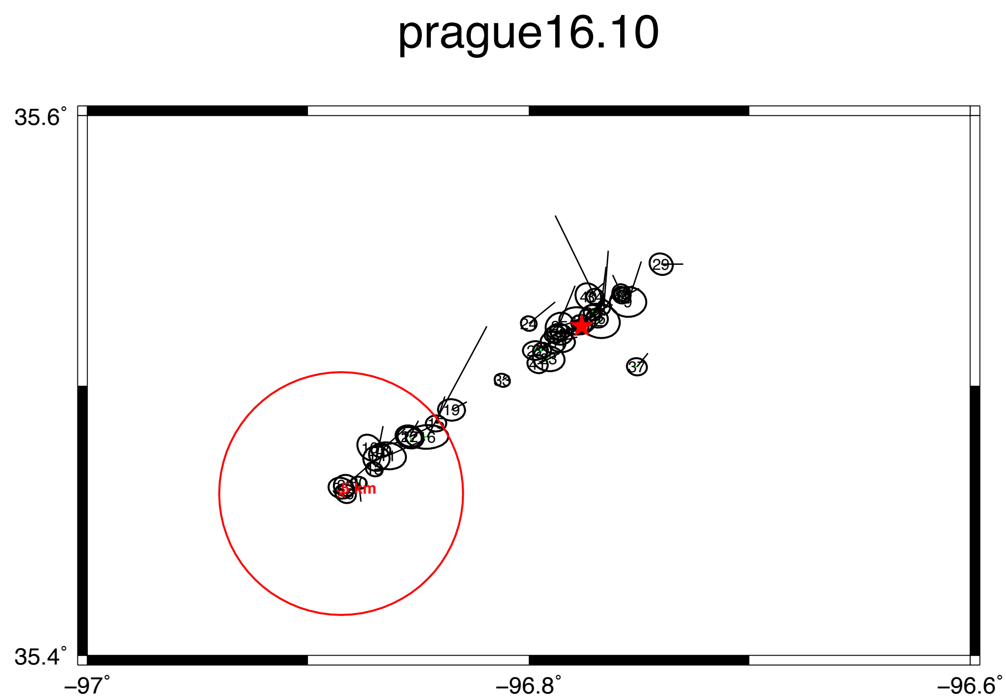

Figure D13. Calibrated locations of the earthquakes in the Prague, Oklahoma cluster. See Figure D1 for further explanation.

Sikeston, Missouri

The sikeston cluster is in the New Madrid Seismic Zone, just south of St. Louis. It consists of 27 events, going back to 2000. Although there are many older events that can be well-located, the connectivity with more recent events is weak, causing convergence problems. The recent events, especially with coverage by the TA, have the best depth control, so I built the cluster with them. The largest magnitude is 4.3. I worked with a subset that could be done with free depth to refine the velocity structure, and then added in the others. Depths range from 4-16 km, but most are less than 10 km deep. Using data out to 1.2° the calibration level is 0.6 km. All 27 events qualify as CE01.

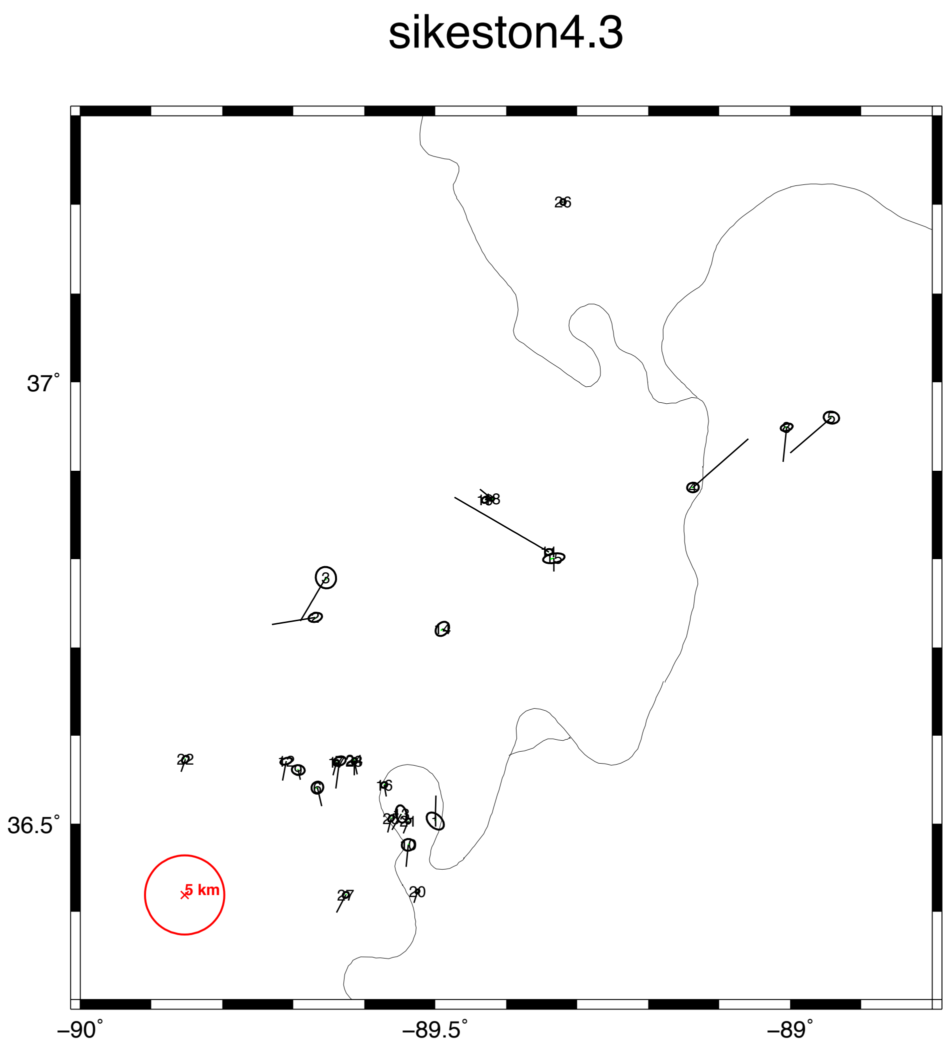

Figure D14. Calibrated locations of the earthquakes in the Sikeston, Missouri cluster. See Figure D1 for further explanation.

Slemp, Kentucky

This cluster is in eastern Kentucky, near the border with Virginia and Tennessee. There is a lot of mining activity in the area and many of the events recorded here are likely to be man-made explosions. It is thought that natural earthquakes also occur here, so resolution of focal depth for this cluster is a major issue.

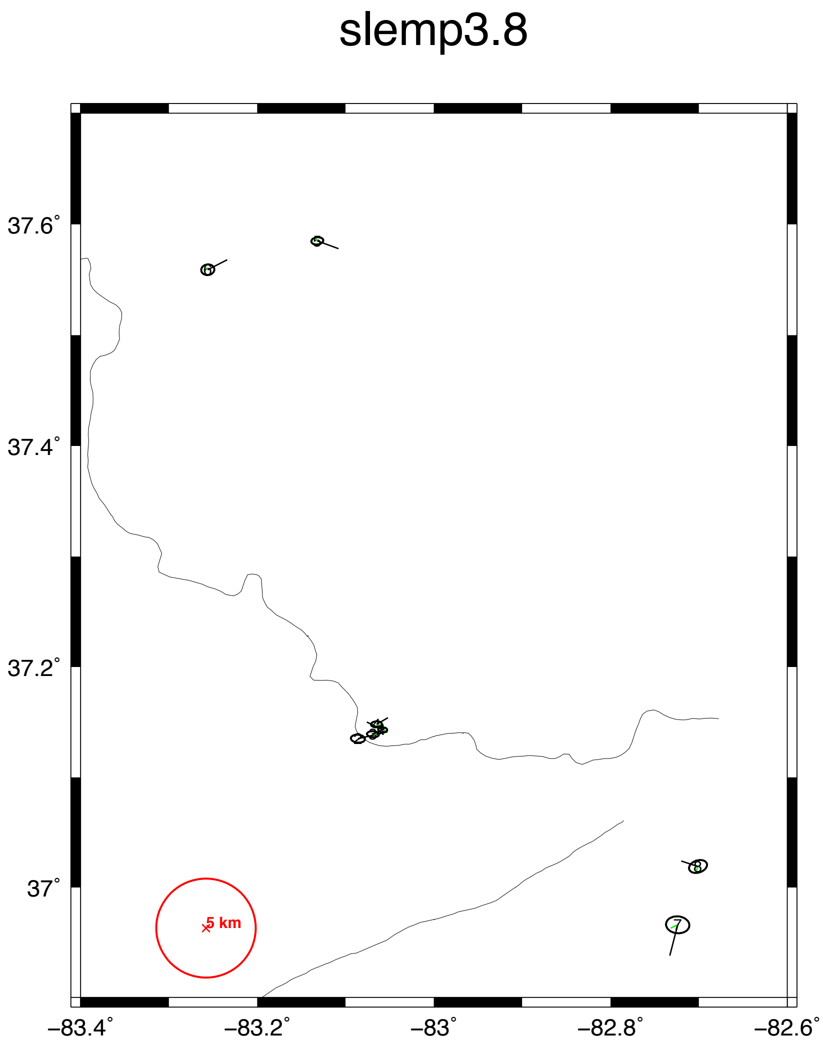

Figure D15. Calibrated locations of the earthquakes in the Slemp, Kentucky cluster. See Figure D1 for further explanation.

I isolated 4 events with close-in readings (<0.1°) and worked with these first. They are a mainshock-aftershock sequence on November 10, 2012. The mainshock is the only event in the area that was observed beyond about 400 km, with 4.2 mb. I refined the velocity structure with these events and obtained very good locations with depths of 20-21 km. I can do a free depth solution, which wants to move them a few km deeper, but that violates the readings at the closest station, so I set them manually. So even though it appears that many of the events in this area are mine blasts, this sequence seems to consist of natural earthquakes. I added four more events with close-in readings to make the final cluster. One of these events (20130129.1657) is still a bit marginal in this respect, with the nearest station about 2 focal depths away. There are many more events that could be added to the cluster but I found that they lacked the data needed to provide a useful constraint on depth: many of them could be held at any depth from the surface to 20 km and yield similar quality locations. The events in this cluster fall into two depth ranges, around 20 (5 events) and around 11 km (3 events). Using data out to 1.2° direct calibration gives a calibration level of 0.6 km and all events qualify as CE01.

Snyder, Texas

The snyder cluster is based on a set of tightly-clustered small events in central Texas near the town of Snyder. 45 events are included in the cluster. It's likely that these are related to injection wells in the area. Magnitudes are up to M4.6. There is very good coverage for direct calibration but there are no stations closer than about 30 km so focal depths cannot be strongly constrained. A depth of 7 km fits well for all events. Using data out to 1.4°, direct calibration gives a calibration level of 0.9 km, and 36 events qualify as CE01-02.

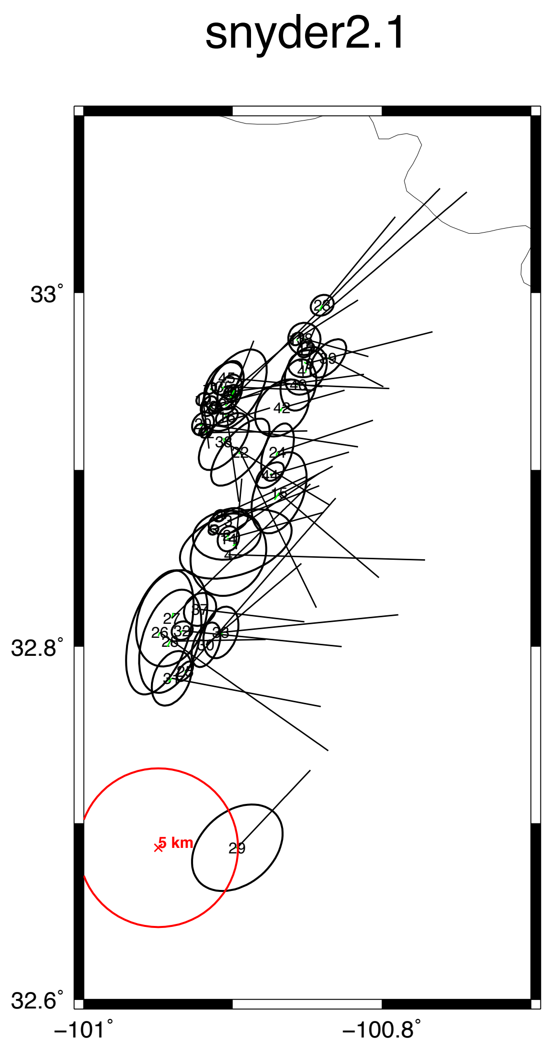

Figure D16. Calibrated locations of the earthquakes in the Snyder, Texas cluster. See Figure D1 for further explanation.

Socorro, New Mexico

This cluster is near Socorro, New Mexico, and represents seismic activity associated with the Rio Grande Rift. It contains 31 events between 1989-2013, ranging up to 4.8 mb. Most of the events have good depth control from stations at close range, but the coverage is not quite adequate for a free-depth relocation. The relocation is done with fixed depths that were set manually. Most events have inferred depths between about 8-13 km but the full range is 4-14 km. Direct calibration is done using readings out to 1.0°, yielding a calibration level of 0.8 km. All events qualify as CE02 or better, and 25 are CE01.

socorro1.14

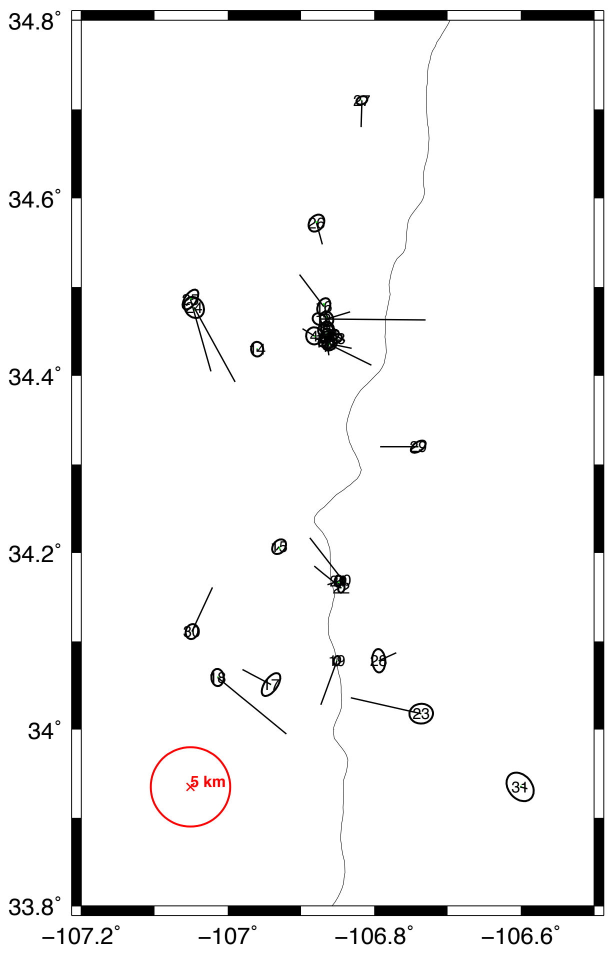

Figure D17. Calibrated locations of the earthquakes in the Soccoro, New Mexico cluster. See Figure D1 for further explanation.

Thurmond, South Carolina

This cluster is named for the Strom Thurmond Reservoir along the Georgia-South Carolina border. It contains only 6 events, but one of them is M4.5. They were well-recorded by the TA. The nearest station is about 20 km, so depth resolution is a little weak but a depth of 8 km works very well for all of them. Using data out to 1.2° direct calibration yields a calibration level of 0.8 km and all 6 events qualify as CE01. There are other events nearby but they have few stations in common with the recent events that were nailed by the TA so the HD method does not work for them.

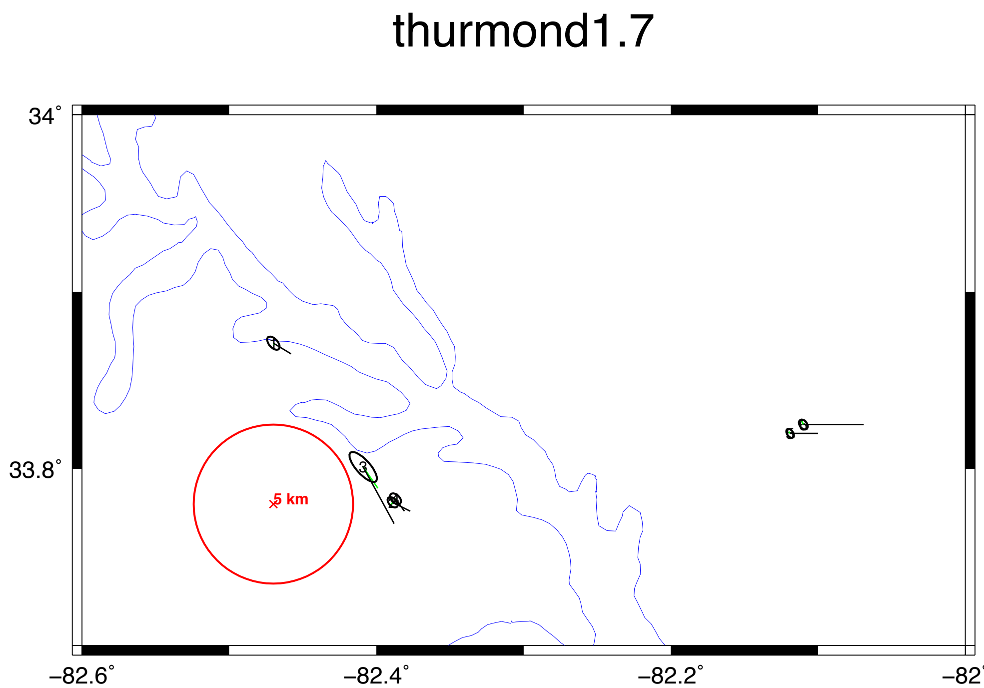

Figure D18. Calibrated locations of the earthquakes in the Thurmond, South Carolina cluster. See Figure D1 for further explanation.

Timpson, Texas

Timpson is a cluster near the Texas-Louisiana border, named for a nearby town. It's small, only 14 events, but several of them were recorded to teleseismic distances. Magnitudes up to 4.5 mb. All events are in the time period 2011-2013. Two of the events had readings very close, less than 10 km, and these were used to set the focal depth and velocity structure. Other events have readings at about 25 km so their depths are fairly well constrained as well. Depths range from 2-11 km, with most in the range 4-9 km. Using data out to 1.4° the azimuthal coverage is very good and a calibration level of 2.1 km is obtained. 12 events qualify as CE02-03. It has been suggested that these events are related to nearby injection wells disposing of waste fluids from hydrofracturing.

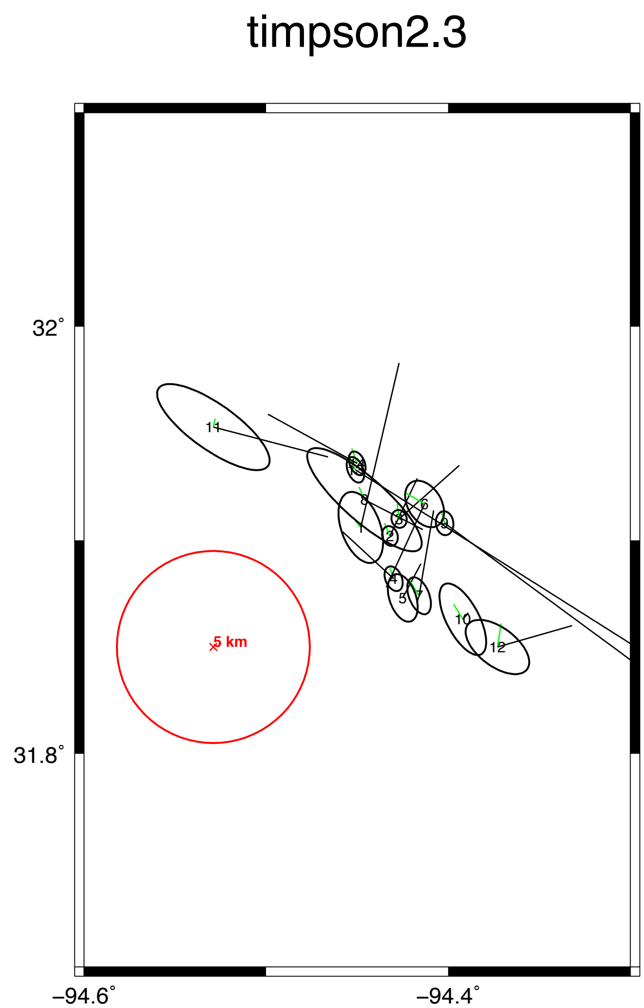

Figure D19. Calibrated locations of the earthquakes in the Timpson, Texas cluster. See Figure D1 for further explanation.

Trinidad, Colorado

The Trinidad cluster is based on the 2011-2012 sequence in southern Colorado (near the city of Trinidad) that included a 5.6 Mw event on August 23, 2011. It includes several earlier events as well, for a total of 63 events. Calibration is possible because of readings from a four-station temporary network deployed by the USGS. 21 events recorded by this network (and also by the permanent stations in the region) were analyzed as a separate cluster, providing free-depth solutions ranging from 4 to 17 km, with uncertainties of 1.1-2.8 km. The other 42 events in the cluster were added in small batches and their depths were adjusted manually to be consistent with the well-constrained events, with estimated uncertainty of 3 km.

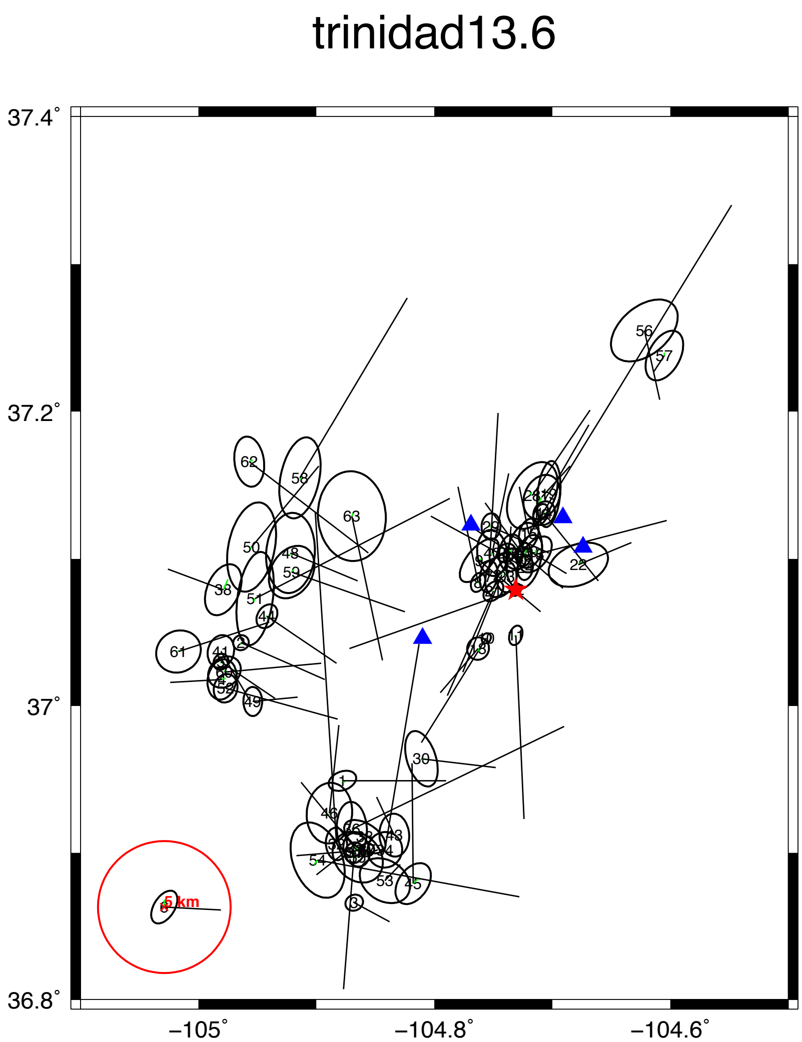

Figure D20. Calibrated locations of the earthquakes in the Trinidad, Colorado cluster. See Figure D1 for further explanation.

Using P and S readings out to 1.0°, the calibration level is 1.3 km and 60 of the 63 events qualify as CE01-CE03 for epicentral uncertainty. Depths range from 4-17 km, with a mean near 10 km.